False discovery rate control with multivariate p-values

Zhiyi Chi[1]

April 20, 2006

Abstract. In multiple hypothesis testing, oftentimes each hypothesis can be assessed by several test statistics, resulting in a multivariate p-value. This raises the question as to how to develop the false discovery rate (FDR) paradigm for multiple testing based on multivariate p-values. On the other hand, a multiple testing procedure based on univariate p-values can have very limited capability of controlling the positive FDR (pFDR) and very low power even when the target FDR control level is moderate. This raises the question as to how to alleviate or overcome the limitation when multiple test statistics are available. To address both questions, we propose and investigate two classes of FDR controlling procedures using multivariate p-values, one incorporating the components of the p-values sequentially, the other incorporating the components simultaneously. Theoretical analysis and simulation study demonstrate that the proposed procedures can improve the pFDR control and power substantially and that, at a given target FDR control level, the improvement depends on the other control parameters of the procedures.

---

[1]Department of Statistics, University of Connecticut, Storrs, CT 06269. Correspondence: zchi@stat.uconn.edu. Partially supported by NIH Grants MH68028-01 and DC007206-01A1.

[2]AMS 2000 subject classification. Primary 62G10, 62H15. Secondary 62G20.

[3]The author thanks Prof. Zhiqiang Tan for helpful suggestions.

1 Introduction

Multiple hypothesis testing in most cases involves multidimensional observations. Sometimes, it suffices to use a single, or “univariate”, statistic per hypothesis for the testing [3, 4, 8, 11–13, 16, 18, 20–25]. However, it is often the case that one needs to apply several statistics to each hypothesis in order to get a reasonable result. For instance, to identify a certain type of sounds from an acoustic signal, it is the norm to examine multiple features of each segment of the signal in order to determine whether or not it contains the type of sound [15]. The features can be evaluated by a vector of marginal or conditional p-values, or a multivariate p-value. A general question is, when multivariate p-values are available, how to use them to control the Type I errors for multiple testing. Under the framework of the FDR control, the article proposes two classes of procedures to address this question.

There are two important aspects of the FDR control, namely the pFDR and power. Recall that the FDR is defined as $E[\frac{V}{R\vee 1}]$ [3], and pFDR as $E[\frac{V}{R} \mid R > 0]$, where R is the number of rejected nulls, and V the number of rejected true nulls, or false discoveries. The two error rates are related by FDR = pFDR$\times P(R > 0)$ [21]. The random fraction $\frac{V}{R\vee 1}$ is usually referred to as the false discovery proportion (FDP) [11]. If there are a total of n nulls and N of them are true, then the (empirical) power is defined as $\frac{R-V}{(n-N)\vee 1}$.

To motivate, consider the following example. Suppose the means of bivariate normal distributions $N(\boldsymbol{\mu}_i, \Sigma_i)$, $i = 1, \ldots, n$, are of interest, with Σ_i being unknown. For each i, the null hypothesis is $H_i : \boldsymbol{\mu}_i = \mathbf{0}$. Suppose it is known that Σ_i is diagonal when H_i is true. To test H_i, a sample of $\nu+1$ iid observations $(X_{ij}, Y_{ij}) \sim N(\boldsymbol{\mu}_i, \Sigma_i)$ are collected. Let $t_{X,i}$ and $t_{Y,i}$ be the t-statistics of $X_{i1}, \ldots, X_{i,\nu+1}$ and $Y_{i1}, \ldots, Y_{i,\nu+1}$, respectively and ξ_{i1} and ξ_{i2} their marginal p-values. How to use the bivariate p-values (ξ_{11}, ξ_{12}), (ξ_{21}, ξ_{22}), …, (ξ_{n1}, ξ_{n2}) to test $H_1, \ldots, H_n$ in order to attain a desired FDR?

For this particular problem, the t-statistics are a reasonable choice not only because of their well established utilities in testing on means but also because of their simplicity. In general, when hypotheses involve relatively complex distributions, it is often desirable to exploit multiple simple statistics. On the one hand, simple test statistics are easy to grasp. On the other, when combined appropriately, simple test statistics can yield satisfactory results for hypothesis testing (cf. [2]).

Continuing the example, suppose it is known that if H_i is false, then both coordinates of $\boldsymbol{\mu}_i$ are positive. Let ξ_{i1} be the upper-tail p-values of $t_{X,i}$ and ξ_{i2} those of $t_{Y,i}$. Does it suffice to only use $\xi_{11}, \ldots, \xi_{n1}$? As is well known [3], the FDR can be controlled at any desired level by using $\xi_{11}, \ldots, \xi_{n1}$ alone. However, this way of testing can cause quite strong limitation on the pFDR control and power. For instance, suppose that, unknown to the investigator, every Σ_i is $\text{diag}(1, 1)$ whether or not H_i is true, and when H_i is false, $\boldsymbol{\mu}_i = (.5, .4)$. Let the procedure of [3], henceforth referred to as the BH procedure, be applied to ξ_{i1}. Suppose $\nu = 8$ and the samples collected for different H_i are independent. Let the fraction of false nulls among all $H_1, \ldots, H_n$ be .05. It turns out that the pFDR is always greater than or equal to $\beta_* \approx .289$. This is in contrast to the FDR, which can be controlled at any level. Moreover, whether or not the FDR is greater than β_* has a critical influence on the power of the BH procedure. In order for the procedure to have a fixed positive power when $n \gg 1$, the FDR must be strictly greater than β_*. If the FDR is less than β_*, the power of the BH procedure drops to 0 at rate $O_p(1/n)$ and its pFDR converges to β_* [6]. The limitation on the pFDR control and power becomes more severe if $\xi_{12}, \ldots, \xi_{n2}$ alone are used. In that case, the pFDR cannot be less than $\beta_* \approx .447$.

The limitation on the pFDR control and power illustrated in the example is not unique to the BH procedure. It occurs whenever the p-values have bounded densities

[6, 7]; see Section 2 for a brief discussion. The limitation affects not only the pFDR control, but also the control of other types of Type I error rates, such as excessive FDP [7].

For the example, it is sensible to exploit both ξ_{i1} and ξ_{i2} for the testing. Using the procedures proposed later, when both p-values are used, the minimum achievable pFDR is reduced to about .017. Therefore, at least in theory, one can attain (p)FDR just a little more than .017 while still attaining a fixed positive power. Section 5 will give more details on multiple hypothesis testing involving t-statistics.

Why should the pFDR be a concern? Oftentimes, follow-up actions ensue only after *some* discoveries are made. Imagine that in a study, an investigator applies a multiple testing procedure that controls the FDR at .05 and he gets some rejected nulls. Without extra data to repeat the same testing procedure, the investigator cannot know whether the FDR control level is too low. In evaluating the rejected nulls at hand, if the investigator is informed that the pFDR cannot be less than, say, .4, his plan for a follow-up study on these nulls is likely to be different from the plan he would make if he believes the pFDR is about the same as the FDR.

In summary, the motivation for our study is two-fold, first, to develop FDR controlling procedures that utilize multivariate p-values, and second, to alleviate potential limitation on multiple hypothesis testing based on univariate p-values. The main focus of our investigation will be the pFDR control and power. There has been relatively little work on how to improve the pFDR control. On the other hand, there has been quite amount of work on how to improve power, with a major finding being that the BH procedure can be made more powerful by incorporating an estimate of the overall fraction of false nulls [3, 4, 20]. The finding has prompted investigations on its estimation [13, 17, 22].

The rest of the article is organized as follows. In Section 2, we set up the basic framework and notations. The investigation is based on a random effects model [9, 13]. Sections 3 and 4 propose two classes of procedures that combine the components of the multivariate p-values sequentially or simultaneously. For each class, after showing that it can control the FDR, we will study its pFDR control and power, specifically, their dependence on the parameters that control how the components of the multivariate p-values are combined. In Section 5, we consider multiple testing involving t-distributions, F-distributions, and Normal distributions and report on a simulation study. Section 6 concludes with a brief discussion. Proofs of theoretical results are given in Appendix.

2 Preliminaries

2.1 Notations

Denote by K the fixed dimension of each multivariate p-value. For a generic point $\boldsymbol{x} \in \mathbb{R}^K$, denote its coordinates by $x_1, \ldots, x_K$. For $c \in \mathbb{R}$, denote $\boldsymbol{c} = (c, \ldots, c) \in \mathbb{R}^K$. For $\boldsymbol{s}$, $\boldsymbol{t} \in \mathbb{R}^K$, denote $\boldsymbol{s} \le \boldsymbol{t}$ if $s_k \le t_k$ for all k and $\boldsymbol{s} < \boldsymbol{t}$ if $\boldsymbol{s} \le \boldsymbol{t}$ and $\boldsymbol{s} \ne \boldsymbol{t}$.

Given null hypotheses $H_1, \ldots, H_n$, the multivariate p-value associated with H_i will be denoted by $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iK})$, with each ξ_{ik} the p-value under a marginal or conditional distribution of the corresponding null distribution. Denote $\theta_i = \mathbf{1}\{H_i \text{ is true}\}$ and a the population fraction of false nulls in the tested hypotheses. We assume that the p-values are generated from the following random effects model [9, 13]:

$$\theta_i \sim \text{Bernoulli}(a), \quad \boldsymbol{\xi}_i \,|\, \theta_i \sim \begin{cases} \text{Unif}(0,1)^{\otimes K} & \text{if } \theta_i = 0 \\ G \text{ with density } g & \text{if } \theta_i = 1 \end{cases} \tag{2.1}$$

where $\boldsymbol{\xi}_i \,|\, \theta_i = 0 \sim \text{Unif}(0,1)^{\otimes K}$ means that given $\theta_i = 0$, $\xi_{i1}, \ldots, \xi_{iK} \overset{\text{iid}}{\sim} \text{Unif}(0,1)$.

Under this model, the joint distribution function of $\boldsymbol{\xi}$ is

$$P(\boldsymbol{\xi} \le \boldsymbol{x}) = (1-a)\prod_{k=1}^{K} x_k + aG(\boldsymbol{x}), \quad \boldsymbol{x} \in [0,1]^K$$

with density $1 - a + ag(\boldsymbol{x})$.

Some remarks are in order. First, when H_i is false, $\xi_{i1}, \ldots, \xi_{iK}$ need not be independent of each other. Multivariate p-values can arise from multidimensional data. For example, for the null $H : \boldsymbol{X} = (X_1, \ldots, X_K) \sim \mathbb{P}$, ξ_k can be the conditional or marginal p-value of X_k under $\mathbb{P}$. More generally, given measurable functions $\phi_1, \ldots, \phi_K$, ξ_k can be chosen to be the conditional or marginal p-value of $\phi_k(\boldsymbol{X})$ under $\mathbb{P}$.

Second, the assumption that conditioning on $\theta = 0$, $\xi_1, \ldots, \xi_K$ are iid $\sim \text{Unif}(0,1)$ can be realized by transformations with conditional distribution functions. For example, following the first remark, let $\mathbb{P}$ have a joint density. Then the conditional distributions

$$F_k(x \,|\, x_1, \ldots, x_{k-1}) = \mathbb{P}(X_k \le x \,|\, X_s = x_s,\ s < k), \quad k = 1, \ldots, K$$

are continuous. If the conditional p-values $\xi_k = F_k(X_k \,|\, X_1, \ldots, X_{k-1})$ are chosen as the components of the multivariate p-value associated with H, then they are iid $\sim \text{Unif}(0,1)$ when H true.

Third, notice that $F_k(X_k \,|\, X_1, \ldots, X_{k-1})$ are themselves functions on $\boldsymbol{X}$, such that their values are independent under H. In general, for measurable functions $\phi_1, \ldots, \phi_K$, if $\phi_1(\boldsymbol{X})$, …, $\phi_K(\boldsymbol{X})$ are independent under H, each having a continuous marginal distribution F_k, then the marginal p-values $\xi_k = F_k(\phi_k(\boldsymbol{X}))$ can be chosen as the components of a multivariate p-value. Again, $\xi_1, \ldots, \xi_K$ are iid $\sim \text{Unif}(0,1)$ when H is true.

2.2 The BH procedure and criticality

The BH procedure can be described as follows [22]. Given a target FDR control level $\alpha \in (0,1)$, for p-values $\xi_1, \ldots, \xi_n$, let $R(t) = \#\{i : \xi_i \le t\}$ and

$$\tau = \sup\left\{t \in [0,1] : \frac{t}{\alpha} \le \frac{R(t) \vee 1}{n}\right\}.$$

Then the BH procedure rejects nulls whose ξ_i are no greater than τ.

Under the random effects model (2.1), the FDR actually realized by the BH procedure is $(1-a)\alpha$ [5, 10, 22]. On the other hand, the "local FDRs" associated with the nulls are

$$P(\theta_i = 1 \,|\, \xi_i) = \frac{1-a}{1-a+ag(\xi_i)} \ge \beta_* := \frac{1-a}{1-a+a\sup g}; \tag{2.2}$$

see [9]. It follows that if $\sup g < \infty$, then, unlike the FDR, pFDR $\ge \beta_* > 0$ [7]. This results in a "criticality phenomenon" of the BH procedure. Roughly speaking, if $\alpha < \alpha_* := \beta_*/(1-a)$, then the power of the BH procedure decreases to 0 at rate $O_p(1/n)$; whereas if $\alpha > \alpha_*$, the power of the BH procedure converges to a positive value at rate $O_p(\sqrt{\log\log n/n})$ [6, 11]. More generally, for any multiple testing procedure based on $\xi_1, \ldots, \xi_n$, decreasing the FDR to below β_* only reduces the power to 0 without decrease in the pFDR. The lower bound β_* solely depends on the distribution of the p-values, rather than on specific testing procedures applied to the p-values [7].

Henceforth, the BH procedure is said to be subcritical (resp. supercritical) if its target FDR control level α is greater (resp. less) than α_*. Unless $a = 1$, α_* is strictly greater than the minimum attainable pFDR, i.e. β_*. To summarize, we have

Proposition 1 Suppose that under the random effects model for $\xi_1, \ldots, \xi_n$, G is strictly concave. If the BH procedure is supercritical, then its power is $O_p(1/n)$; whereas if the BH procedure is subcritical, its power is $G(u) + O_p(\sqrt{\log\log n/n})$, with u the (unique) positive solution to $x/\alpha = (1-a)x + aG(x)$.

Likewise, it is not difficult to see that, when using multivariate p-values for multiple testing, the pFDR may still be bounded away from 0, and therefore the limitation on the pFDR control and power in general cannot be completely overcome.

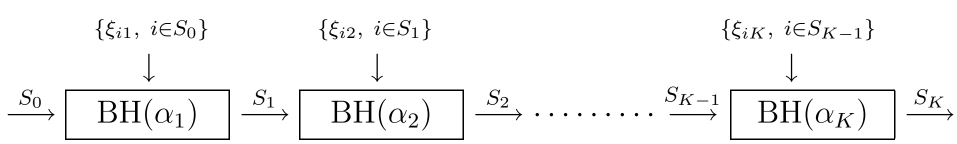
Proposition 2 Under the random effects model for multivariate p-values, for any multiple testing procedure, pFDR $\geq (1-a)\alpha_*$, where α_* is defined as

$$\alpha_* = \frac{1}{1 - a + a \sup g}. \qquad \square \quad (2.3)$$

Nevertheless, Proposition 2 also indicates that by using multivariate p-values, it is possible to alleviate the limitation. Indeed, because $\sup g \geq \sup f$ for any marginal density f of g, a lower pFDR can be attained by using all the components of the p-values than by using part of them. This observation is the basis for our proposed procedures.

3 Sequential combination

3.1 Description and the FDR control



Given the target FDR control level α, fix $\alpha_k \in (0, 1)$, such that $\alpha_1 \cdots \alpha_K = \alpha$. Set $S_0 = \{H_1, \ldots, H_n\}$. As the diagram illustrates, at the beginning, all the hypotheses are treated as being rejected. At step $k \geq 1$, only hypotheses rejected by all previous steps (S_{k-1}) are tested based on the kth component ξ_{ik} with the target FDR control level α_k. Only hypotheses rejected by all the steps are output as rejections. More specifically,

Sequential procedure

1. For $k = 1, \ldots, K$, denote $R_{k-1} = \#S_{k-1}$. Let

$$\tau_k = \sup\left\{t \in [0,1] : \frac{t}{\alpha_k} \le \frac{\#\{H_j \in S_{k-1}:\ \xi_{jk} \le t\} \vee 1}{R_{k-1}}\right\}. \tag{3.1}$$

   Then set $S_k = \{H_j \in S_{k-1}:\ \xi_{jk} \le \tau_k\}$.

2. Only the null hypotheses in S_K are rejected.

In principle, for each step, the BH procedure can be replaced with some other multiple testing procedure. The advantage of using the BH procedure is that it allows the FDR to be computed exactly. We first need to see how the sequential procedure controls the FDR. The result below shows that the fraction of true nulls among the rejected ones keeps decreasing as component p-values are incorporated in sequel.

Theorem 1 For $k \ge 1$, let V_k be the number of true nulls in those that are rejected by the first k steps of the sequential procedure, i.e. $V_k = \#\{H_j \in S_k:\ H_j \text{ is true null}\}$. Recall that a is the population fraction of false nulls. Then FDR $= (1-\xi)\alpha$. Indeed,

$$E\left[\frac{V_k}{R_k \vee 1}\right] = (1-a)\prod_{s=1}^{k} \alpha_s, \quad k = 1, \ldots, K. \qquad \square \tag{3.2}$$

A similar result on $E[V_k/(R_k \vee 1)]$ holds for the sequential procedure in a frequentist setting. That is, if there are N true null hypotheses at the beginning, then with $a = 1 - N/n$, equation (3.2) still holds.

3.2 Dynamics of the sequential procedure

Roughly speaking, there are two reasons why the sequential procedure can attain better pFDR control. First, in each step of the procedure, the target FDR control level can be chosen relatively high to keep the BH procedure in that step subcritical. Second, after each step, the fraction of true nulls among all the rejected nulls is decreased. The reduced fraction of true nulls lowers the critical value for the BH procedure in the next step, which also helps maintain subcriticality; see more detail below. As long as each step is subcritical, the entire procedure has a positive power. At the same time, the overall pFDR can still be low. To see this more clearly, consider the following measures

(1) $V_k/(R_k \vee 1)$: the FDP after the kth step, which is also the fraction of true nulls among the nulls tested by the $(k+1)$th step;

(2) $R_k/(R_{k-1} \vee 1)$: the fraction of rejected nulls retained by the kth step;

(3) $(R_k - V_k)/(n - V_0)$: the compound empirical power of the first k steps; and

(4) $(R_k - V_k)/[(R_{k-1} - V_{k-1}) \vee 1]$: the relative power of the kth step.

Given a target FDR control level α, a fraction of false nulls a, and a distribution function F on $[0, 1]$, define

$$\tau(\alpha, a, F) = \sup\left\{t : \frac{t}{\alpha} \le (1-a)t + aF(t)\right\} \tag{3.3}$$

We say that $G(\boldsymbol{s})$ is argument-wise strictly concave, if for each k and fixed values of s_j, $j \ne k$, $G(\boldsymbol{s})$ is strictly concave in s_k.

Theorem 2 Under the random effects model (2.1), suppose G is argument-wise strictly concave and has a continuous density. Let $a_0 = a$ and for $k \ge 1$,

$$a_k = 1 - E\left[\frac{V_k}{R_k \vee 1}\right] = 1 - (1-a)\prod_{s=1}^{k} \alpha_k\,.$$

Denote by G_1 the marginal distribution of ξ_1 under G and suppose that

$$u_k := \tau(\alpha_k,\, a_{k-1},\, G_k) > 0, \quad k \ge 1, \tag{3.4}$$

where G_k is inductively defined as

$$G_k(x) = G\left(\xi_k \le x \,|\, \xi_1 \le u_1, \ldots, \xi_{k-1} \le u_{k-1}\right), \quad k \ge 2. \tag{3.5}$$

Let $R_0 = n$ and V_0 be the total number of true nulls. Then as $n \to \infty$, $R_k \xrightarrow{\text{a.s.}} \infty$, and

$$\frac{V_k}{R_k} \xrightarrow{\text{a.s.}} 1 - a_k = (1-a)\prod_{s=1}^{k} \alpha_k, \quad \frac{R_k}{R_{k-1}} \xrightarrow{\text{a.s.}} \frac{u_k}{\alpha_k},$$
$$\frac{R_k - V_k}{n - V_0} \xrightarrow{\text{a.s.}} \prod_{s=1}^{k} G_s(u_s) = \frac{u_1 \cdots u_k a_k}{\alpha_1 \cdots \alpha_k a}, \tag{3.6}$$
$$\frac{R_k - V_k}{R_{k-1} - V_{k-1}} \xrightarrow{\text{a.s.}} G_k(u_k) = \frac{u_k a_k}{\alpha_k a_{k-1}}.$$

By dominated convergence, the pFDR of the entire procedure tends to α. □

Basically, Theorem 2 says that as long as each step of the sequential procedure is subcritical, the fraction of true nulls $V_k/R_k \approx 1 - a_k$ will keep dropping step by step, and the rates of rejections as well as the powers will all stabilize at positive values. Eq. (3.4) gives the condition for each step being subcritical. Eq. (3.5) characterizes how the distributions of the component p-values ξ_k associated with false nulls are conditioned by the sequential rejections.

To see why reducing the fraction of true nulls V_k/R_k helps maintain subcriticality, suppose that for each false null, the component p-values $\xi_1, \ldots, \xi_K$ are independent. Then each G_k in (3.5) is a concave marginal distribution with density g_k. If the BH procedure is directly applied to the kth component p-values associated with *all* the nulls, the critical value for its target FDR control level is $\alpha_* = \{1 + a[g_k(0) - 1]\}^{-1}$. Whereas in the sequential procedure, the corresponding critical value is $\alpha'_* = \{1 + a_k[g_k(0) - 1]\}^{-1}$. Since $\xi_k > \xi$ and $g_k(0) > 1$, $\alpha'_* < \alpha_*$, and hence the BH procedure can attain a lower pFDR with a positive power when it is used in the sequential procedure.

The following corollary will be used to analyze the power of the sequential procedure.

Corollary 1 For $u_1, \ldots, u_K > 0$ in (3.4), they consist a solution to

$$\rho(\alpha) \prod_{k=1}^{K} u_k = G(\boldsymbol{u}), \quad \text{with} \;\; \rho(\alpha) = \frac{1}{a}\left(\frac{1}{\alpha} - 1 + a\right) \tag{3.7}$$

and the asymptotic power of the sequential procedure is $\rho(\alpha)\prod_{k=1}^{K} u_k$.

3.3 The pFDR control and power of the sequential procedure

By Proposition 2, the sequential procedure is still constrained by a critical value α_* for the target FDR control level α. The questions are (1) when $\alpha > \alpha_*$, can the sequential procedure attain pFDR $\le \alpha$; and (2) how to maximize its power at the same time?

For the sequential procedure, the pFDR and power depend on the target FDR control levels for individual steps, i.e., $\alpha_1, \ldots, \alpha_K$. Theorem 3 shows that when $\alpha > \alpha_*$, it is possible to select appropriate values of $\alpha_1, \ldots, \alpha_K$ to attain pFDR $\le \alpha$ with a positive power. However, the result gives no indication on how to do this.

Theorem 3 Assume the same conditions for G as in Theorem 2. Given $\alpha \in (\alpha_*, 1)$, there are $\alpha_1, \ldots, \alpha_K \in [0, 1]$ with $\alpha = \alpha_1 \cdots \alpha_K$, such that the sequential procedure with target FDR control levels α_k for individual steps attains pFDR $= (1-a)\alpha + o(1)$ with a positive power, as $n \to \infty$. □

In Theorem 3, it is possible that $\alpha_k = 1$ for some k. For each such k, except for a $o(1)$ fraction, all the nulls previously rejected are still rejected by the kth step. As a result, the kth component p-values are virtually useless and hence can be ignored.

What is the maximum power that the sequential procedure can attain at a given target FDR control level $\alpha \in (\alpha_*, 1)$? Can the maximum power be attained no matter in what order the component p-values are incorporated? From a computational point of view, it is desirable to have an affirmative answer to the second question. We have

Theorem 4 Suppose G is argument-wise strictly concave and has a continuous density g with $g(\mathbf{0}) = \sup g > 1$. Let $\alpha \in (\alpha_*, 1)$. If the target FDR control level of the entire sequential procedure is α, then its power is upper-bounded by

$$P_*(\alpha) = \sup \left\{ G(\boldsymbol{u}) \ : \ \boldsymbol{u} \in [0,1]^K \text{ is a solution to (3.7)} \right\} ,$$

which can be attained with appropriate target FDR controls levels $\alpha_1, \ldots, \alpha_K$ for individual steps. Furthermore, for any permutation $j_1, \ldots, j_K$ of 1, ..., K, $P_*(\alpha)$ can be attained if the procedure applies to ξ_{i,j_1}, ..., ξ_{i,j_K} in sequel. □

3.4 Power of the sequential procedure using p-values with an unbounded density

By Theorem 3, if the p-values have an unbounded density, then in principle the sequential procedure can attain arbitrarily small pFDR. The power of the procedure nevertheless still depends on the control parameters. This point will be illustrated by two examples involving bivariate p-values, showing that the maximum power may or may not be attained by incorporating both components of the p-values. This is in contrast to a later result (Proposition 8), which shows that under mild conditions, for multivariate p-values with a bounded density, all their components must be incorporated in order to attain a low enough pFDR with a positive power.

Example 1 Let $\boldsymbol{X}_n = (X_{n1}, X_{n2})$, $n \geq 1$, be independent, such that if H_i is true, then $\boldsymbol{X}_i \sim N(0, I)$, and otherwise $\boldsymbol{X}_i - \boldsymbol{\mu} \sim N(0, I)$, where $\mu_k > 0$. Let $\boldsymbol{\xi}_i$ consist of $\xi_{ik} = 1 - \Phi(X_{ik})$, $k = 1, 2$, where $\Phi(x)$ is the distribution function of $N(0, 1)$. If H_i is false, then for $k = 1, 2$, the distribution function of ξ_{ik} is

$$\begin{aligned} G_k(u) &= P\left(1 - \Phi(X_{ik}) \le u\right) = P\left(X_{ik} \ge \Phi^*(1-u)\right) \\ &= 1 - \Phi(\Phi^*(1-u) - \mu_k) = \Phi(\Phi^*(u) + \mu_k)\,, \\ \implies \quad G_k'(u) &= \frac{\Phi'(y + \mu_k)}{\Phi'(y)} = \exp\left\{-\mu_k y - \mu_k^2/2\right\}, \quad y = \Phi^*(u)\,. \end{aligned} \tag{3.8}$$

Since $\mu_k > 0$, G_k is strictly concave with $G_k'(0) = \infty$, so the sequential procedure based on $\boldsymbol{\xi}_n$ can attain arbitrarily small pFDR. The next result implies that, at pFDR $\ll 1$, the power is maximized only when both $\alpha_k < 1$, i.e., both ξ_{ik} are incorporated.

Proposition 3 Denote $L(t) = -\log(at)$. Let $\hat{\alpha}_1 = \hat{\alpha}_1(\alpha)$ be the target FDR control level for the first step when the sequential procedure attains pFDR $= (1-a)\alpha$ with the maximum power. Then as $\alpha \to 0$,

$$L(\hat{\alpha}_1) \sim \frac{L(\alpha)\mu_1^2}{\mu_1^2 + \mu_2^2}$$

with the maximal power

$$P_*(\alpha) \sim \frac{1}{\sqrt{a\alpha}} \exp\left\{-\frac{\mu_1^2 + \mu_2^2}{8} - \frac{L(\alpha)^2}{2(\mu_1^2 + \mu_2^2)}\right\}. \tag{3.9}$$

Example 2 Denote by $\text{Exp}(\mu)$ the exponential distribution with mean μ. Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ are iid such that if H_i is true, then $\boldsymbol{X}_i \sim \text{Exp}(1)^{\otimes 2}$ and otherwise $\boldsymbol{X}_i \sim \text{Exp}(\mu_1) \otimes \text{Exp}(\mu_2)$, where $\mu_k > 1$. Let ξ_{ik} be the marginal upper-tail p-values of X_{ik}. Then $\xi_{ik} = e^{-X_{ik}}$ and has distribution function $G_k(u) = u^{1/\mu_k}$, which is strictly concave with $G_k'(0) = \infty$. The next result shows that in general, the maximum power cannot be attained when both ξ_{i1} and ξ_{i2} are incorporated by the sequential procedure.

Proposition 4 Let $\mu_1, \mu_2 > 1$. Then at the target FDR control level $\alpha \in (0, 1)$, the maximum attainable power is $P_*(\alpha) = \rho(\alpha)^{-1/c}$, where $\rho(\alpha)$ is defined in (3.7) and

$c = \max\{\mu_1, \mu_2\} - 1$. If $\mu_1 = \mu_2$, then $P_*(\alpha)$ is attained for any $\alpha_1, \alpha_2 \in [0,1]$ with $\alpha_1\alpha_2 = \alpha$. If $\mu_1 \neq \mu_2$, and μ_k is the larger one, then $P_*(\alpha)$ is attained only when $\alpha_k = \alpha$ and the other α_j is 1.

4 Simultaneous combination

4.1 Description and the FDR control

The simultaneous procedure draws an idea from [22], which treats the end point of the random rejection interval for the p-values in the BH procedure as a stopping time of a martingale running backward in time. The generalization to multivariate p-values is illustrated by Fig. 4.1. The p-values are regarded as points in $[0,1]^K$. A suitable martingale can be constructed running backward along a path $\boldsymbol{\gamma}$ in $[0,1]^K$. Then all the p-values dominated under the partial order $\le$ by a stopping point $\boldsymbol{t}_0 = \boldsymbol{\gamma}(\tau)$ are rejected.

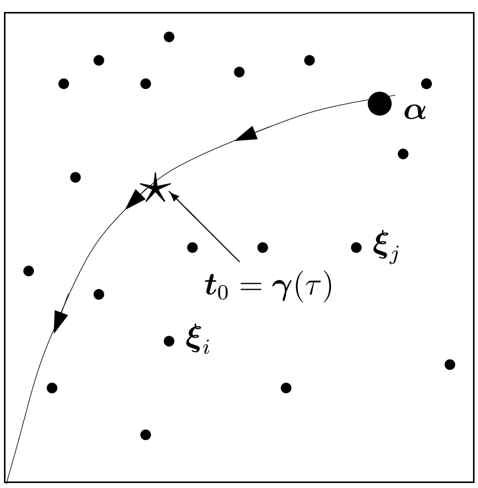


Figure 4.1. An illustration for the simultaneous procedure. The path $\boldsymbol{\gamma}(t)$ is defined on $[0,1]$ and is monotone in the sense that $\boldsymbol{\gamma}(s) < \boldsymbol{\gamma}(t)$ when $s < t$. $\boldsymbol{\alpha} = \boldsymbol{\gamma}(1)$ is the vector of the FDR control parameters. In generate, $\boldsymbol{\gamma}(0)$ need not be $\mathbf{0}$.

Suppose $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ are the multivariate p-values associated with $H_1, \ldots, H_n$. Let α

be the target FDR control level. For $\boldsymbol{t} \in [0,1]^K$, define

$$R(\boldsymbol{t}) = \#\{i = 1, \ldots, n : \boldsymbol{\xi}_i \le \boldsymbol{t}\}, \\ V(\boldsymbol{t}) = \#\{i = 1, \ldots, n : \boldsymbol{\xi}_i \le \boldsymbol{t},\ H_i \text{ is true}\}, \tag{4.1}$$

Let $f_1, \ldots, f_K$ be continuous non-decreasing functions on $[0,1]$ such that

$$0 \le f_k(t) \le 1, \quad f_1(t) \cdots f_K(t) = t, \quad t \in [0,1]. \tag{4.2}$$

Fix a vector of FDR control parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K) \in [0,1]^K$, such that $\prod_{k=1}^K \alpha_k = \alpha$. Then define path $\boldsymbol{\gamma} : [0,1] \to [0,1]^K$ by

$$\boldsymbol{\gamma}(t) = (\alpha_1 f_1(t), \ldots, \alpha_K f_K(t)). \tag{4.3}$$

The above conditions imply that $f_1(1) = \cdots = f_K(1) = 1$ and $\alpha_k \ge \alpha$. They also allow $f_k(0) > 0$ or $\alpha_k = 1$. Especially, if $f_k(t) \equiv 1$ and $\alpha_k = 1$, then from the description below, the kth component p-values $\xi_{1k}, \ldots, \xi_{nk}$ are essentially not used by the simultaneous procedure.

There are several equivalent descriptions of the procedure. The analysis will be based on the following one.

Simultaneous procedure. Define

$$\tau = \sup\left\{s \in [0,1] : s \le \frac{R(\boldsymbol{\gamma}(s)) \vee 1}{n}\right\} \tag{4.4}$$

Reject H_i if $\boldsymbol{\xi}_i \le \boldsymbol{\gamma}(\tau)$, $i = 1, \ldots, n$. □

For $K = 1$, $\boldsymbol{\gamma}(t) = \alpha t$. Then the procedure rejects H_i if $\xi_i \le \boldsymbol{\gamma}(\tau) = \tilde{\tau}/\alpha$, where

$$\tilde{\tau} = \sup\left\{s \in [0,1] : \frac{s}{\alpha} \le \frac{R(s) \vee 1}{n}\right\}.$$

Comparing with the description in [22], it can be seen that in this case, the simultaneous procedure is identical to the BH-procedure.

Although the procedure allows the functions $f_1, \ldots, f_K$ to be chosen quite freely, it will be seen that in order to attain the maximum power, it suffices to use power functions $f_k(t) = t^{a_k}$ with suitable $a_k \geq 0$ (cf. Theorem 7).

For numerical computation, the next description is more suitable. It follows from the observation that $R(\boldsymbol{\gamma}(s))$ has a jump at s if and only if there is i, such that $\boldsymbol{\xi}_i \leq \boldsymbol{\gamma}(s)$ but for any $t < s$, $\boldsymbol{\xi}_i \not\leq \boldsymbol{\gamma}(t)$.

Simultaneous procedure, description 2.

1. For each $i = 1, \ldots, n$, compute $s_i = \max\{k = 1, \ldots, K : f_k^*(\xi_{ik}/\alpha_i)\}$, where

$$f_k^*(x) = \begin{cases} \inf\{s \in [0,1] : x \leq f_k(s)\} & \text{if } x \leq 1 \\ \infty & \text{otherwise} \end{cases}$$

2. Sort $s_1, \ldots, s_n$ into $s_{(1)} \leq s_{(2)} \leq \cdots \leq s_{(n)}$. Define $s_{(0)} = 0$ and set

$$l = \max\left\{k \geq 0 : s_{(k)} \leq \frac{\alpha k}{n}\right\} \tag{4.5}$$

3. Reject H_i if $s_i \leq s_{(l)}$, $i = 1, \ldots, n$. □

The next description does not involve stopping time and is formulated as a more straightforward multidimensional generalization of the BH-procedure.

Simultaneous procedure, description 3. Define

$$D = \left\{\boldsymbol{t} \in [0,1]^K : t_k \leq \alpha_k f_k\left(\frac{R(\boldsymbol{t}) \vee 1}{n}\right),\ k = 1, \ldots, K\right\}. \tag{4.6}$$

Let $\boldsymbol{t}_0 = \sup D$ in the partial order $\leq$. Reject H_i if $\boldsymbol{\xi}_i \leq \boldsymbol{t}_0$, $i = 1, \ldots, n$. □

In general, under the partial order $\leq$, there is no guarantee that the supremum of a set is a singleton. However, $\sup D$ is indeed a singleton and hence $\boldsymbol{t}_0$ is well-defined.

Proposition 5 The set $\sup D$ has only one element $\boldsymbol{t}_0 = \boldsymbol{\gamma}(\tau)$ and the three descriptions are equivalent. □

For the procedure in (4.4), the numbers of rejections and false rejections are

$$R = \sum_{i=1}^{n} \mathbf{1}\left\{\boldsymbol{\xi}_i \le \boldsymbol{\gamma}(\tau)\right\}, \quad V = \sum_{i=1}^{n} (1-\theta_i)\mathbf{1}\left\{\boldsymbol{\xi}_i \le \boldsymbol{\gamma}(\tau)\right\},$$

respectively, where τ is defined in (4.4). Then we have

Theorem 5 For the FDR control of the simultaneous procedure, $E\left[\frac{V}{R\vee 1}\right] = (1-a)\alpha$.

4.2 Criticality and power for the simultaneous procedure

In this section, suppose $g(\mathbf{0}) = \sup g > 1$. By Proposition 2, the pFDR is lower bounded by $(1-a)\alpha_*$, where $\alpha_* = 1/(1-a+ag(\mathbf{0}))$. The value of $g(\mathbf{0})$ can be ∞. Assume that g as a function taking values in $[0,\infty]$ is continuous on $[0,1]^K$.

The pFDR control and power of the simultaneous procedure can be understood by utilizing the random variable τ in (4.4). The main result on τ is that it asymptotically has a fixed point similar to the one characterized by [6, 11]. Define

$$h(s) = (1-a)\alpha s + aG(\boldsymbol{\gamma}(s)). \tag{4.7}$$

Then $h(s)$ is continuous with $h(0) = 0$ and $h(1) = (1-a)\alpha < 1$. As a result, the set $\{s \in [0,1] : s = h(s)\} \ne \emptyset$. Additionally, all solutions to $s = h(s)$ are strictly less than 1.

Proposition 6 Suppose $g(\mathbf{0}) = \sup g > 1$. If

$$s^* = \sup\{s \in [0,1] : s = h(s)\} > 0, \tag{4.8}$$

then as $n \to \infty$, $\tau \xrightarrow{\text{a.s.}} s^*$. In particular, (4.8) holds if

$$\begin{aligned} &\lim_{\boldsymbol{x}\to\mathbf{0}} g(\boldsymbol{x}) = g(\mathbf{0}); \\ &f_k(0) = 0,\ k = 1, \ldots, K; \text{ and} \\ &f_1, \ldots, f_K \text{ are strictly increasing around } 0. \end{aligned} \tag{4.9}$$

Theorem 6 Given $\alpha \in (\alpha_*, 1)$, if (4.8) is satisfied, then as $n \to \infty$, $R \xrightarrow{\text{a.s.}} \infty$, and

$$\begin{cases} \dfrac{V}{R} \xrightarrow{\text{a.s.}} (1-a)\alpha \\ \dfrac{R-V}{n-V_0} \xrightarrow{\text{a.s.}} G(\boldsymbol{\gamma}(s^*)) = \dfrac{s^*}{a}\left[1-(1-a)\alpha\right] . \end{cases} \tag{4.10}$$

By dominated convergence, the pFDR of the simultaneous procedure tends to α.

In particular, under condition (4.9), (4.8) holds and so the above conclusions hold.

□

Theorem 6 implies that if the target FDR control level α is strictly greater than α_*, then, comparing with the sequential procedure, it is easier for the simultaneous procedure to attain pFDR $= (1-a)\alpha + o(1)$ with a positive power. First, it is easy to find $f_1, \ldots, f_K$ satisfying the conditions in (4.9), and hence ensure (4.8). For example, we can choose $f_k(x) = x^{q_k}$, with $q_k \in (0,1)$, $\sum_{k=1}^K q_k = 1$. Then (4.10) holds for any $\boldsymbol{\alpha} \in [0,1]^K$ with $\prod_{k=1}^K \alpha_k = \alpha$, yielding pFDR $= (1-a)\alpha + o(1)$ and power > 0. Whereas for the sequential procedure, in each step, one needs to carefully ensure subcriticality in order for the entire procedure to attain pFDR $\approx (1-a)\alpha$ with a positive power.

Let $\boldsymbol{u} = \boldsymbol{\gamma}(s^*)$. Then $0 < u_k \le \alpha_k \le 1$ and $u_1 \cdots u_K = s^*\alpha$. The next corollary to (4.10) provides a useful expression of the power of the simultaneous procedure.

Corollary 2 Under the same condition as in Theorem 6, as $n \to \infty$, the empirical power $(R-V)/(n-V_0) \xrightarrow{\text{a.s.}} G(\boldsymbol{u})$, with $\boldsymbol{u}$ a solution to

$$\rho(\alpha)\prod_{k=1}^K u_k = G(\boldsymbol{u}), \quad \text{with} \quad \rho(\alpha) = \frac{1}{a}\left(\frac{1}{\alpha} - 1 + a\right). \qquad \square \tag{4.11}$$

The next result gives the maximum power that can be attained by the simultaneous procedure. It is similar to the one for the sequential procedure but requires weaker assumptions on G; cf. Theorem 4.

Theorem 7 Let $\alpha \in (\alpha_*, 1)$. Then the power of the simultaneous procedure is upper-bounded by $0 < P_*(\alpha) = \sup\{G(\boldsymbol{u}) : \boldsymbol{u} \in D\} < 1$, where

$$D = \left\{\boldsymbol{u} \in [0,1]^K : \boldsymbol{u} \text{ is a solution to (4.11)}\right\}.$$

The upper-bound can be attained with appropriate selections of $\boldsymbol{\alpha}$ and $f_1, \ldots, f_K$. Furthermore, it suffices to choose from functions $f_k(t) = t^{q_k}$, $k = 1, \ldots, K$ with $q_k \geq 0$ and $q_1 + \cdots + q_K = 1$. □

It is possible that in the selections of $\boldsymbol{\alpha}$ and $f_1, \ldots, f_K$ that yield the maximum power $P_*(\alpha)$, $\alpha_k = 1$ and $f_k(t) \equiv 1$ for some of the k's. For each such k, the components $\xi_{1k}, \ldots, \xi_{nk}$ of the p-values are essentially not used by the procedure.

4.3 When using multivariate p-values is better than using univariate p-values?

Section 3.4 gives an example showing that when the nulls are on the means of bivariate exponential distributions, it yields the maximum power to only use the p-values associated with the variate that has the larger mean. The example raises the general question as to when using multivariate p-values may increase the power. The next result gives a criteria on when multivariate p-values cannot improve the power.

Proposition 7 Let G have a continuous density g with $g(\mathbf{0}) = \sup g$. Suppose

$$G(\boldsymbol{x}) < \max\{G_1(|\boldsymbol{x}|), \ \ldots, \ G_K(|\boldsymbol{x}|)\}, \text{ if } x_k < 1 \text{ for at least two } k, \tag{4.12}$$

where $|\boldsymbol{x}| = \prod_{k=1}^K x_k$ and G_k is the kth marginal distribution function of G. Then for any $\alpha \in (\alpha_*, 1)$, there is k and a BH procedure testing on $\xi_{1k}, \ldots, \xi_{nk}$ with the target FDR control level α, such that it has a higher power than both the sequential and the simultaneous procedures testing on $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ with the target FDR control level α. □

For the example in Section 3.4, by $G_k(x) = x^{a_k}$, $k = 1, 2$, with $a_k = 1/\mu_k$, it is seen that $G_1(x_1)G_2(x_2) < (x_1 x_2)^{\min\{a_1,a_2\}} = \max\{G_1(|\boldsymbol{x}|), G_2(|\boldsymbol{x}|)\}$. Therefore, the example is a special case of Proposition 7. From (4.12), $G(t, \ldots, t)/t^K \le \max_k \{G_k(t^K)/t^K\}$. Let $t \to 0$ to get $g(\mathbf{0}) \le \max\{g_k(0)\}$, with g_k the marginal density of the kth component of the multivariate p-value. By $g(\mathbf{0}) = \sup g$, either $g(\mathbf{0}) = \infty$ or $\mathbf{0}$ is not the unique maximum point of g. On the other hand, when $g(\boldsymbol{x}) < g(\mathbf{0}) < \infty$ for all $\boldsymbol{x} \ne \mathbf{0}$, the next result implies that for any sequential or simultaneous procedure which attains pFDR $\approx (1 - a)\alpha_*$ with a positive power, a null is rejected only when all the components of its p-value are small enough. Thus, all the components have to be incorporated.

Proposition 8 Suppose $g(\boldsymbol{x}) < g(\mathbf{0}) < \infty$ for any $\boldsymbol{x} \ne \mathbf{0}$. Then, as $\alpha \to \alpha_*$, for any sequence of solutions $\boldsymbol{u}$ to $\rho(\alpha)\prod_{k=1}^K u_k = G(\boldsymbol{u})$, the limit is $\mathbf{0}$.

5 Numerical results

This section reports a simulation study on the procedures proposed in previous sections. Throughout, the multivariate p-values are of $K = 2$ dimensions. In each simulation, the distributions under true nulls and those under false nulls are of the same type but have different parameter values. The distributions involved in the simulations are t-, F-, or Normal distributions. The nulls are sampled under the random effects model with the fraction of false nulls $a = .02$ or $.05$.

Each simulation has a single target FDR control level α and consists of 1500 runs. In each run, a set of $n = 10,000$ bivariate p-values are sampled and then tested with procedures with different control parameters. The pFDR and the power are computed as Monte Carlo averages from all the runs and then examined as functions of the control parameters. For the sequential procedure, the target FDR control level $\alpha_1 \in [\alpha, 1]$ for the first step is taken as the free parameter. In the simultaneous procedure, $f_1(x) = x^q$ and $f_2(x) = x^{1-q}$ with $q \in [0, 1]$. For each fixed $q = s/5$, $s = 0, 1, \ldots, 5$, the pFDR and the power are examined as functions of α_1. All the simulations are conducted using R language [19].

The following fact is used to evaluate the minimum attainable pFDR. Suppose the null is $H : X \sim F_0$. Then the upper-tail p-value $1 - F_0(X)$ of $X \sim F$ has distribution function $G(u) = F(F_0^{-1}(1-u))$ and therefore has density

$$G'(u) = \frac{F'(\phi(u))}{F_0'(\phi(u))}, \quad \text{with } \phi(u) = F_0^{-1}(1-u). \tag{5.1}$$

5.1 Multiple testing involving *t*-distributions

Introduction gives an example of multiple testing on t-statistics and asserts that the pFDR has a positive lower bound. To verify the claim, recall the density of the noncentral t-distribution with ν degrees of freedom (df) and noncentrality parameter δ is

$$t_{\nu,\delta}(x) = \frac{\nu^{\nu/2}}{\sqrt{\pi}\,\Gamma(\nu/2)} \frac{e^{-\delta^2/2}}{(\nu + x^2)^{(\nu+1)/2}} \sum_{k=0}^{\infty} \Gamma\left(\frac{\nu+k+1}{2}\right) \frac{(\delta x)^k}{k!} \left(\frac{2}{\nu + x^2}\right)^{k/2}.$$

Denote $t_\nu(x) = t_{\nu,0}(x)$. Then

$$\frac{t_{\nu,\delta}(x)}{t_\nu(x)} = e^{-\delta^2/2} \sum_{k=0}^{\infty} \Gamma\left(\frac{\nu+k+1}{2}\right) \frac{(\delta x)^k}{k!} \left(\frac{2}{\nu + x^2}\right)^{k/2} \Big/ \Gamma\left(\frac{\nu+1}{2}\right). \tag{5.2}$$

Let $\delta > 0$. Then $t_{\nu,\delta}(x)/t_\nu(x)$ is strictly increasing on $(0, \infty)$ and

$$\lim_{x\to\infty} \frac{t_{\nu,\delta}(x)}{t_\nu(x)} = e^{-\delta^2/2} \sum_{k=0}^{\infty} \Gamma\left(\frac{\nu+k+1}{2}\right) \frac{(\sqrt{2}\,\delta)^k}{k!} \Big/ \Gamma\left(\frac{\nu+1}{2}\right) < \infty. \tag{5.3}$$

Under the setting of the example in Introduction, suppose that if H_i is true, then $\boldsymbol{\mu}_i = \mathbf{0}$ and $\Sigma_i = \Sigma_0$ is diagonal, and otherwise $\boldsymbol{\mu}_i = (c_1, c_2)$ and $\Sigma_i = \Sigma_a = (\sigma_{jk})$, where

$c_i > 0$. Then for false H_i, $t_{X,i} \sim t_{\nu,\delta}$ with $\delta = \sqrt{\nu+1}c_1/\sigma_{11}$, and by (5.1)–(5.3), the distribution function G_1 of ξ_{i1} is strictly concave on $[0, \frac{1}{2}]$, with $G_1'(u) > G_1'(1-u)$ for $u \le \frac{1}{2}$ and $\sup G_1'(x) = G_1'(0) < \infty$. By Proposition 1, the minimum pFDR attainable by using $\xi_{11}, \ldots, \xi_{n1}$ alone is $\beta_* = (1-a)/(1-a+aG_1'(0))$. Now, as in the example, let $a = .05$, $\nu = 8$, $(c_1, c_2) = (.5, .4)$, and $\Sigma_0 = \Sigma_a = \text{diag}(1, 1)$. Then $G_1'(0) \approx 46.81$ and $\beta_* \approx .289$. Likewise, let G_2 be the distribution function of ξ_{i2} when H_i is false. Then $G_2'(0) \approx 23.47$ and the corresponding $\beta_* \approx 0.447$. On the other hand, the joint density of $\boldsymbol{\xi}_i$ when H_i is false is $G_1'(x)G_2'(y)$. The minimum attainable pFDR by using both ξ_{i1} and ξ_{i2} is $(1-a)/(1-a+aG_1'(0)G_2'(0)) \approx .017$. The claims of the example is thus verified.

In the numerical study, $\Sigma_0 \equiv \text{diag}(1, 1)$, $(c_1, c_2) \equiv (.75, .7)$, and $\mu \equiv 6$. We conduct six simulations, one for each combination of $a \in \{.05, .02\}$ and $\Sigma_a \in \left\{I, AA^T, BB^T\right\}$, where I is the identify matrix and

$$A = \begin{pmatrix} 1 & .2 \\ .3 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -.2 \\ -.3 & 1 \end{pmatrix}.$$

The target FDR control levels α in the simulations are collected in Table 5.1. Also collected are $\alpha_*^{(k)} = 1/(1-a+aG_k'(0))$, the critical values of the FDR control level for the BH procedure using ξ_{i1} or ξ_{i2} alone. The target FDR control level α are set equal to $\hat{\alpha}_1\hat{\alpha}_2$ for some $\hat{\alpha}_k > \alpha_*^{(k)}$. Note $\alpha < \alpha_*^{(k)}$ in all the simulations.

The top panel of Fig. 5.1 displays the power as a function of α_1. Although the plots are not directly comparable as they are associated with different α, they have similar shapes. The maximum powers and corresponding values of α_1 are reported as $(P_*^{(SE)}, \hat{\alpha}_1)$ in Table 5.1. The bottom panel of Fig. 5.1 shows how the FDR and the pFDR depend on α_1. In all the plots, the α_1-FDR curve is basically flat, showing that the sequential procedure can control the FDR at level $(1-a)\alpha$ no matter the value of $\alpha_1 \in [\alpha, 1]$. In contrast, when α_1 is close to α or to 1, the pFDR is substantially greater than the FDR, indicating the sequential procedure becomes supercritical. The substantial increase in pFDR is accompanied with the substantial decrease in power.

The power of the simultaneous procedure is also strongly affected by the control parameters. The top panel of Fig. 5.2 displays the power as a function of α_1 for different $f_1(x) = x^q$, with $(a, \Sigma_a) \equiv (.05, AA^T)$. The maximum powers and corresponding values of α_1 are collected in Table 5.1. The bottom panel of Fig. 5.2 plots the (p)FDR as functions of α_1. For q near .5 ($q = .4, .6$), the two functions are identical on $\alpha_1 \in [\alpha, 1]$. This is observed for all the combinations of a and Σ_a in Table 5.1. The result is consistent with Theorem 6, which implies that the simultaneous procedure is subcritical as long as $q \neq 0, 1$. However, when q is near 0 or 1 ($q = .2, .8$), the pFDR exhibits some difference from the FDR, indicating that the number of p-values ($n = 10,000$) is not yet large enough for the asymptotic result of Theorem 6 to fully take effect. The difference between the pFDR and the FDR is greater for $a = .02$ (results not shown). In the extreme case $q = 0$, $f_1(x) \equiv 1$ and $f_2(x) = x$. When $\alpha_1 = 1$, ξ_{i1} essentially are not incorporated and the simultaneous procedure is equivalent to the BH procedure applied to ξ_{i2} alone with target FDR control level α. As a result, the simultaneous procedure is supercritical, yielding a large gap between the pFDR and the FDR. As α_1 decreases to α, the gap decreases. In Fig. 5.2, the gap vanishes when $\alpha_1 = \alpha$. However, in general, the gap can stay positive for all $\alpha_1 \geq \alpha$. Similar comments apply to the case $q = 1$ as well.

From Table 5.1, it is seen that a simultaneous procedure can attain approximately the same maximum power as the sequential procedure, provided that $f_1(x) = x^q$ is selected appropriately. For $a = .05$, q can be chosen in a wider range ($.2 \leq q \leq .8$) to attain a maximum power comparable to that of the sequential procedure. When

Table 5.1: Simulation results involving t-distributions. $(P_*^{(SE)}, \hat{\alpha}_1)$ is the maximum power of the sequential procedure and the corresponding value of α_1. $(P_{*,q}^{(SI)}, \hat{\alpha}_1)$ is the maximum power of the simultaneous procedure with power functions $f_1(x) = x^q$ and $f_2(x) = x^{1-q}$, and the corresponding value of α_1. $\Sigma_1 = AA^T$, $\Sigma_2 = BB^T$

	$(.05, I)$	$(.02, I)$	$(.05, \Sigma_1)$	$(.02, \Sigma_1)$	$(.05, \Sigma_2)$	$(.02, \Sigma_2)$
α	.0756	.24	.1	.2912	.1	.2912
$\alpha_*^{(1)}$	.2387	.4395	.2639	.4727	.2639	.4727
$\alpha_*^{(2)}$	.2837	.4976	.3437	.5669	.3437	.5669
$(P_*^{(SE)}, \hat{\alpha}_1)$	(.112, .54)	(.184, .78)	(.247, .62)	(.308, .82)	(.049, .52)	(.112, .78)
$(P_{*,\,0}^{(SI)}, \hat{\alpha}_1)$	(.052, .076)	(.033, .24)	(.114, .1)	(.040, .291)	(.018, .1)	(.022, .291)
$(P_{*,\,0.2}^{(SI)}, \hat{\alpha}_1)$	(.109, .076)	(.141, .24)	(.242, .1)	(.225, .291)	(.044, .1)	(.075, .291)
$(P_{*,\,0.4}^{(SI)}, \hat{\alpha}_1)$	(.114, .14)	(.192, .26)	(.25, .18)	(.325, .291)	(.051, .14)	(.115, .291)
$(P_{*,\,0.6}^{(SI)}, \hat{\alpha}_1)$	(.114, .4)	(.193, .72)	(.25, .42)	(.325, .72)	(.051, .42)	(.116, .84)
$(P_{*,\,0.8}^{(SI)}, \hat{\alpha}_1)$	(.113, 1)	(.161, 1)	(.250, .98)	(.273, 1)	(.049, .98)	(.097, 1)
$(P_{*,\,1}^{(SI)}, \hat{\alpha}_1)$	(.069, 1)	(.050, 1)	(.17, 1)	(.07, 1)	(.026, 1)	(.035, 1)

$a = .02$, only the maximum powers associated with $q \in \{.4, .6\}$ are about the same as that of the sequential procedure. The fraction of false nulls a also affects how sensitive the power of the simultaneous procedure is to α_1. In Fig. 5.3, with $q = .6$, the α_1-power curve is plotted for the six combinations of a and Σ_a. The plots in the top row are associated with $a = .05$, while those in the bottom row with $a = .02$. The latter ones are much flatter and only exhibit weak modalities, but have higher maximum values.

5.2 Multiple testing involving $\boldsymbol{F}$-distributions

Recall that for $p, q = 1, 2, \ldots$ and $d \ge 0$, the distribution of

$$\frac{\sum_{i=1}^{p}(Z_i + \mu_i)^2/p}{\sum_{j=1}^{q} \tilde{Z}_j^2/q}$$

is called noncentral F-distribution with (p, q) df and noncentrality parameter d and denoted by $F_{p,q,d}$, where $Z_1, \ldots, Z_p$, $\tilde{Z}_1, \ldots, \tilde{Z}_q \overset{\text{iid}}{\sim} N(0,1)$ and $\mu_1^2 + \cdots + \mu_p^2 = d$. Let $\rho = p/q$. The density of $F_{p,q,d}$ is

$$f_{p,q,\delta}(x) = e^{-\delta/2}\rho^{p/2}x^{q/2-1}(1+\rho x)^{(p+q)/2}\sum_{k=0}^{\infty}\frac{(\delta/2)^k}{k!\,B\left(\frac{p}{2}+k,\ \frac{q}{2}\right)}\left(\frac{\rho x}{1+\rho x}\right)^k, \quad x \ge 0.$$

Denote by $f_{p,q}(x)$ the density of $F_{p,q,d}$ when $d = 0$. Then

$$\frac{f_{p,q,\delta}(x)}{f_{p,q}(x)} = e^{-\delta/2}B\left(\frac{p}{2}, \frac{q}{2}\right)\sum_{k=0}^{\infty}\frac{(\delta/2)^k}{k!\,B\left(\frac{p}{2}+k,\ \frac{q}{2}\right)}\left(\frac{\rho x}{1+\rho x}\right)^k, \tag{5.4}$$

which is strictly increasing. Furthermore

$$\lim_{x\to\infty}\frac{f_{p,q,\delta}(x)}{f_{p,q}(x)} = e^{-\delta/2}B\left(\frac{p}{2}, \frac{q}{2}\right)\sum_{k=0}^{\infty}\frac{(\delta/2)^k}{k!\,B\left(\frac{p}{2}+k,\ \frac{q}{2}\right)} < \infty. \tag{5.5}$$

In our simulations involving F-distributions, the bivariate p-values are sampled as follows. For each $i = 1, \ldots, n$, a pair of independent random variables $X_i \sim F_{p_1,q_1,d_i}$ and $Y_i \sim F_{p_2,q_2,d_2}$ are drawn. If H_i is true, then $d_1 = d_2 = 0$; otherwise, $d_1 = \delta_1 > 0$ and $d_2 = \delta_2 > 0$, where δ_k are fixed. The corresponding bivariate p-value is $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2})$, where ξ_{i1} and ξ_{i2} are the marginal p-values of X_i and Y_i under F_{p_1,q_1} and F_{p_2,q_2}, respectively. By (5.1), (5.4), and (5.5), for $k = 1, 2$, the distribution function G_k of ξ_{ik} is strictly concave and

$$\sup G_k'(x) = G_k'(0) = \lim_{x\to\infty}\frac{f_{p_k,q_k,\delta_k}(x)}{f_{p_k,q_k}(x)} < \infty.$$

We conduct two simulations. In both simulations, we set $(p_1, q_1, \delta_1) = (5, 5, 10)$, $(p_2, q_2, \delta_2) = (7, 6, 10)$. The values of a and α in the two simulations are given in

Table 5.2: Simulation results involving F-distributions. The meanings of the entries are similar to those in Table 5.1.

	$a = .05$	$a = .02$
α	.2475	.5184
$\alpha_*^{(1)}$	.4239	.6478
$\alpha_*^{(2)}$	.5298	.7381
$(P_*^{(SE)}, \hat{\alpha}_1)$	(.083, .72)	(.128, .88)
$(P_{*,\,0}^{(SI)}, \hat{\alpha}_1)$	(.014, .2475)	(.018, .5184)
$(P_{*,\,0.2}^{(SI)}, \hat{\alpha}_1)$	(.060, .2475)	(.078, .5184)
$(P_{*,\,0.4}^{(SI)}, \hat{\alpha}_1)$	(.087, .2475)	(.135, .52)
$(P_{*,\,0.6}^{(SI)}, \hat{\alpha}_1)$	(.087, .62)	(.141, 1)
$(P_{*,\,0.8}^{(SI)}, \hat{\alpha}_1)$	(.077, 1)	(.103, 1)
$(P_{*,\,1}^{(SI)}, \hat{\alpha}_1)$	(.027, 1)	(.029, 1)

Table 5.2. The values of $\alpha_*^{(k)} = 1/(1-a+aG_k'(0))$ are also listed, which are the critical values for the target FDR control level of the BH procedure testing on ξ_{ik} alone. The target FDR control level α are set equal to $\hat{\alpha}_1\hat{\alpha}_2$ for some $\hat{\alpha}_k > \alpha_*^{(k)}$. Similar to the results for t-distributions, for the sequential procedures, the plots of power *vs* α_1 are unimodal, and the plots of FDR *vs* α_1 and those of pFDR *vs* α_1 are different when the power is low (Fig. 5.4). For the simultaneous procedure, the plots of power *vs* α_1 exhibit no or very week modality, even though for appropriately selected $f_1(x) = x^q$, the procedure is subcritical for all $\alpha_1 \in [\alpha, 1]$; see the plots for $q = .4$, and .6 in Fig. 5.5.

5.3 Multiple testing involving Normal distributions

Following (3.8), it can be shown that for $\eta \sim N(\mu, \sigma^2)$, the upper-tail p-value under $H : \eta \sim N(0, 1)$ has density

$$\exp\left\{-\frac{1}{2\sigma^2}\left((1-\sigma^2)y^2 + \mu y + \mu^2\right)\right\}, \quad y = \Phi^*(u). \tag{5.6}$$

Therefore, if $\sigma = 1$ and $\mu > 0$, then the distribution function of the p-value is strictly concave with $g(0) = \infty$; whereas if $\sigma < 1$, then the distribution function is strictly convex in a neighborhood of 0 with $g(0) = 0$.

In our simulations involving Normal distributions, the bivariate p-values are sampled as follows. For each $i = 1, \ldots, n$, $(X_i, Y_i) \sim N(\boldsymbol{\mu}, \Sigma)$ is drawn. If H_i is true, then $\boldsymbol{\mu} = \mathbf{0}$, $\Sigma = I$; otherwise $\mu_k = d_k > 0$, $k = 1, 2$, and $\Sigma = (\sigma_{ij})$ with $\sigma_{11} = \sigma_{22} = 1$ and $\sigma_{12} = \sigma_{21} = \rho \in (-1, 1)$. Clearly, ρ is the correlation coefficient of X_i and Y_i. The corresponding bivariate p-value $\boldsymbol{\xi}_i$ consists of the marginal upper-tail p-values of X_i and Y_i under $N(0, 1)$, i.e., $\xi_{i1} = 1 - \Phi(X_i)$, $\xi_{i2} = 1 - \Phi(Y_i)$.

If H_i is false, then from (5.6), the distribution of ξ_{i1} is strictly concave with density ∞ at 0. On the other hand, conditioning on $X_i = x$, $Y_i \sim N(\mu_2 + \rho(x - \mu_1), 1 - \rho^2)$. By (5.6), unless $\rho = 0$, the conditional distribution of ξ_{i2} given ξ_{i1} is not concave. Furthermore, if $\rho < 0$, then for $x - \mu_1$ large, $E(Y_i \,|\, X_i = x) < 0$. Consequently, for $\xi_{i1} \ll 1$, the conditional distribution of ξ_{i2} is strictly convex. Because the BH procedure only rejects nulls with small p-values and has 0 power asymptotically when the distribution function of the p-values is convex, it follows that by incorporating both ξ_{i1} and ξ_{i2}, the sequential procedure may become less capable of rejecting false nulls.

To numerically examine the effect of ρ, in the simulations, $\boldsymbol{\mu} = (1.5, 1.5)$, $\rho \in \{0, \pm.5, \pm.9\}$, and $a \in \{.02, .05\}$. The target FDR control level is fixed at .15. We only show the results for $a = .05$. The results for $a = .02$ are qualitatively similar.

For the sequential procedure, when $\rho = 0$, both steps of the procedure are subcritical, so it is not surprising that the pFDR and the FDR are close to each other for all the values of α_1; see Fig. 5.6, bottom (1). The α_1-power curve exhibits a similar shape as the curves from the simulations involving t- or F-distributions; see Fig. 5.6, top (1). It is worth noticing that the curve is asymmetric on the log-scale of α_1; see plot (1′). This shows that in order to increase the power at the desired pFDR level, the first step of the sequential procedure should be quite "generous" by setting the target FDR control level relatively high, and let the second step further reduce the fraction of false rejections.

When $\rho > 0$, in spite of the fact that the conditional distribution of ξ_{i2} given ξ_{i1} is not concave around 0, for $\alpha = .15$, which is moderately small, the sequential procedure behaves similarly as in the case $\rho = 0$; see plots (2)–(3). On the other hand, when $\rho < 0$, the sequential procedure behaves significantly differently. For moderate negative correlation ($\rho = -.5$), the power decreases moderately when both ξ_{i1} and ξ_{i2} are incorporated. The pFDR also deviates from FDR moderately. However, when there is a strong negative correlation ($\rho = -.9$), ξ_{i1} and ξ_{i2} can be described as completely working against each other. When both are incorporated, the power drops to 0 and pFDR becomes 1, and so the procedure completely loses the capability of detecting false nulls.

For the simultaneous procedures, similar results are obtained. Briefly, for $\rho \geq 0$, by incorporating both ξ_{i1} and ξ_{i2}, the power is increased and the pFDR is controlled. Whereas for $\rho < 0$, the power is decreased when both ξ_{i1} and ξ_{i2} are incorporated. When $\rho = -.9$, incorporating both ξ_{i1} and ξ_{i2} causes the power to drop to 0 and the pFDR to increase to 1. For brevity, plots for the results are omitted.

6 Discussion

Scientific or engineering exploratory studies often involve multiple testing. In many problems, such as pattern detection for images or acoustic signals, multiple test statistics are often used to assess each individual hypothesis, yielding a multivariate p-value. This is especially the case when the data exhibits multiple facets, each containing some unique information that can be exploited for the identification of subtle but important patterns from noisy background.

The current FDR paradigm for multiple testing has been focused on the case where each individual hypothesis is evaluated with a single, or univariate, p-value. As shown here and elsewhere [6, 7], multiple testing procedures based on univariate p-values can suffer some severe limitation on their power and capability of controlling the pFDR. As a result, they cannot achieve a satisfactory trade-off between the Type I error rate and the power, as characterized by the so-called "ROC curve". The limitation is in contrast to the extraordinary ROC performance of biological visual systems, which are capable of incorporating multiple levels of features extracted from inputs [14]. In order to broaden the applicability of the FDR paradigm, it is therefore necessary to explore how to utilize multivariate statistics for multiple testing, when such statistics are available.

In this work, we propose and investigate two classes of FDR controlling procedures that combine multivariate p-values. The sequential procedure acts like a series of BH procedures which "filter" out true nulls step by step. The simultaneous procedure transforms each multivariate p-value into one number, and then, in the same fashion as the BH procedure, identifies interesting nulls based on the derived numbers.

Our theoretical and numerical studies demonstrate that under certain conditions, the power and the pFDR control can be improved by incorporating multivariate p-values. The improvement relies on the control parameters. To illustrate this point, we use Monte Carlo simulations and exhaustive search to identify values of the control parameters that maximize the power. In the presence of a large number of tests, it can be shown that the maximization is about the same as maximizing the number of rejected nulls [6]. In practice, one only has a single set of p-values. Exhaustive search to maximize the number of rejections may potentially lead to an incorrect selection of parameter values in a fashion similar to overfitting. How to choose parameter values appropriately based on a single data set needs to be addressed in the future.

Our study also shows another aspect of multiple testing using multivariate p-values. That is, in not chosen appropriately, the component p-values associated with different test statistics can interact to negatively impact the power and the pFDR control. How to select appropriate test statistics in order to effectively separate false nulls from true nulls is a problem of "feature selection", which has been a challenging issue [1]. On the other hand, when the test statistics are already selected, there may still be some ways to avoid the situation as in the example in Section 5.3, where two negatively correlated test statistics completely rid the sequential procedure of power. Part of the problem is due to the nature of the BH procedure, which only rejects nulls with p-values less than a threshold. One may instead reject nulls with p-values within certain random intervals, as suggested in [6]. Another possibility is to derive new test statistics based on the available ones. For the example in Section 5.3, under a false null, $\frac{1}{\sqrt{2}}(X_i + Y_i)$ and $\frac{1}{\sqrt{2}}(-X_i + Y_i))$ has mean values $1.5\sqrt{2}$ and 0, respectively. The marginal p-values of these derived random variables have less negative interaction with each other and so may serve better for the sequential procedure. In general, the transformation of available test statistics into more useful ones is a challenging problem as well.

References

[1] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Comput.*, 9(7):1545–1588, 1997.

[2] Y. Amit and D. Geman. A computational model for visual selection. *Neural Comput.*, 11(7):1691–1715, 1999.

[3] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995.

[4] Y. Benjamini and Y. Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educational & Behavioral Statistics*, 25(1):60–83, 2000.

[5] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.*, 29(4):1165–1188, 2001.

[6] Z. Chi. Criticality of a false discovery rate control procedure. Technical report, University of Connecticut, Department of Statistics, 2005.

[7] Z. Chi and Z. Tan. Positive false discovery proportions for multiple testing: intrinsic bounds and adaptive control. Technical report, University of Connecticut and the John Hopkins University, Department of Statistics, 2005.

[8] S. Dudoit, M. J. van der Laan, and K. S. Pollard. Multiple testing. I. Single-step procedures for control of general type I error rates. *Stat. Appl. Genet. Mol. Biol.*, 3:Art. 13, 71 pp. (electronic), 2004.

[9] B. Efron, R. Tibshirani, J. D. Storey, and V. G. Tusher. Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, 96(456):1151–1160, 2001.

[10] H. Finner and M. Roters. On the false discovery rate and expected type I errors. *Biometri. J.*, 43:985–1005, 2001.

[11] C. Genovese and L. Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *J. Roy. Statist. Soc. Ser. B*, 64(3):499–517, 2002.

[12] C. Genovese and L. Wasserman. Exceedance control of the false discovery proportion. Technical Report 807, Carnegie Mellon University, Department of Statistics, 2004.

[13] C. Genovese and L. Wasserman. A stochastic process approach to false discovery control. *Ann. Stat.*, 32(3):1035–1061, 2004.

[14] Y. Jin and S. Geman. Context and hierarchy in a probabilistic image model. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, New York, 2006.

[15] J. A. Kogan and D. Margoliash. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparison study. *J. Acoust. Soc. Am.*, 103(4):2185–2196, 1998.

[16] E. L. Lehmann and J. P. Romano. Generalizations of the familywise error rate. *Ann. Stat.*, 33(3):1138–1154, 2005.

[17] N. Meinshausen and J. Rice. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Stat.*, 34(1):in press, 2006.

[18] M. Perone Pacifico, C. Genovese, I. Verdinelli, and L. Wasserman. False discovery control for random fields. *J. Amer. Statist. Assoc.*, 99(468):1002–1014, 2004.

[19] R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0, URL http://www.R-project.org.

[20] J. D. Storey. A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B*, 64(3):479–498, 2002.

[21] J. D. Storey. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Stat.*, 31(6):2012–2035, 2003.

[22] J. D. Storey, J. E. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. Roy. Statist. Soc. Ser. B*, 66(1):187–205, 2004.

[23] M. J. van der Laan, S. Dudoit, and K. S. Pollard. Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat. Appl. Genet. Mol. Biol.*, 3:Art. 15, 27 pp. (electronic), 2004.

[24] M. J. van der Laan, S. Dudoit, and K. S. Pollard. Multiple testing. II. Step-down procedures for control of the family-wise error rate. *Stat. Appl. Genet. Mol. Biol.*, 3:Art. 14, 35 pp. (electronic), 2004.

[25] M. J. van der Laan, S. Dudoit, and K. S. Pollard. Multiple testing. Part III. Procedures for control of the generalized family-wise error rate and proportion of false positives. Technical Report 141, UC Berkeley, Division of Biostatistics, 2004.

Appendix: Proofs of theoretical results

A.1 Technical details for the sequential procedure

Proof of Theorem 1. For $k \geq 1$, if $V_{k-1} = 0$, then it is clear that

$$E\left[\frac{V_k}{R_k \vee 1} \,\middle|\, V_{k-1}, R_{k-1}\right] = 0 = \frac{\alpha_k V_{k-1}}{R_{k-1} \vee 1}.$$

Now suppose $V_{k-1} > 0$. Then $R_{k-1} > 0$. By assumption, when a null is true, the components of its p-value are iid $\sim$ Unif$(0,1)$. Therefore, conditioning on that a true null H_i has been rejected by the first $k-1$ steps, $\xi_{ik} \sim$ Unif$(0,1)$ and is independent of ξ_{jk} associated with the other nulls rejected by the first $k-1$ steps. Notice that the independence may not hold if H_i is a false null. Based on the observation, by [22],

$$E\left[\frac{V_k}{R_k \vee 1} \,\middle|\, V_{k-1}, R_{k-1}\right] = \frac{\alpha_k V_{k-1}}{R_{k-1}}\,.$$

As a result, no matter whether V_{k-1} is zero,

$$E\left[\frac{V_k}{R_k \vee 1} \,\middle|\, V_{k-1}, R_{k-1}\right] = \frac{\alpha_k V_{k-1}}{R_{k-1} \vee 1}\,.$$

Taking expectation with respect to (V_{k-1}, R_{k-1}) yields

$$E\left[\frac{V_k}{R_k \vee 1}\right] = \alpha_k E\left[\frac{V_{k-1}}{R_{k-1} \vee 1}\right]\,.$$

By induction and noticing $R_0 = n$ and $EV_0 = (1-a)n$,

$$E\left[\frac{V_k}{R_k \vee 1}\right] = \alpha_1 \ldots \alpha_k E\left[\frac{V_0}{R_0}\right] = \alpha_1 \ldots \alpha_k (1-a)\,.$$

The proof is thus complete. □

To prove Theorem 2, it is necessary to take into account that if a null is false, then the components of its p-value may be dependent. The following way to think about the sequential procedure is helpful. To start with, draw an iid sample $\theta_1, \ldots, \theta_n$ from

Bernoulli(a) and fix it throughout the sequential procedure. For each i, if $\theta_i = 0$, then draw $\xi_{i1} \sim \text{Unif}(0,1)$, otherwise, draw $\xi_{i1} \sim G_1$, where G_1 is the marginal distribution on ξ_1 under the joint distribution G on $(\xi_1, \ldots, \xi_K)$. Apply the BH procedure to all ξ_{i1} and collect the rejected nulls. This finishes the first step. At step $k+1$, for each H_i rejected by all the first k steps, if $\theta_i = 0$, then draw $\xi_{i,k+1} \sim \text{Unif}(0,1)$, otherwise, draw $\xi_{i,k+1}$ from the conditional distribution of ξ_{k+1} given $\xi_1 = \xi_{i1}, \ldots, \xi_{k-1} = \xi_{ik}$. Then apply the BH procedure to the sampled $\xi_{i,k+1}$ and collect the rejected nulls to finish the step.

Given random variables η_i and nonrandom numbers $x_i > 0$, denote $\eta_i = o_p(x_i)$ if $\eta_i / x_i \xrightarrow{\text{a.s.}} 0$ as $k \to \infty$.

Proof of Theorem 2. We show the asymptotics by induction based on the above description. Recall the definition of S_k in (3.1) and u_k in (3.4). For $k = 1$, it is not hard to modify the in-probability result of [11] to get $S_1 = \{H_i : \xi_{i1} \le u_1\} \triangle A_1$ with $|A_1| = o_p(n)$. By $\xi_{i1} \overset{\text{iid}}{\sim} h(x) = (1-a)x + aG_1(x)$ and the Law of Large Numbers (LLN),

$$\frac{R_1}{R_0} = \frac{R_1}{n} \xrightarrow{\text{a.s.}} h(u_1) = \frac{u_1}{\alpha_1}\,.$$

On the other hand, $|R_1 - V_1| = |\{H_i : \theta_i = 1,\ \xi_{i1} \le u_1\}| + o_p(n)$. So by the LLN,

$$\frac{R_1 - V_1}{R_0 - V_0} = \frac{R_1 - V_1}{n - V_0} \xrightarrow{\text{a.s.}} G_1(u_1)\,.$$

From [22], $V_1/R_1 \to (1-a)\alpha_1$. Since $V_0/R_0 \xrightarrow{\text{a.s.}} a$, it follows that

$$\frac{R_1 - V_1}{R_1} \xrightarrow{\text{a.s.}} 1 - (1-a)\alpha_1 = a_1 \implies \frac{V_1}{R_1} \xrightarrow{\text{a.s.}} 1 - (1-a)\alpha_1.$$

It is easy to check that the above convergences imply (3.6). For the induction, we need to justify (3.5) for $k = 2$. The (random) marginal distribution of ξ_{i2} associated with

$H_i \in S_1$ is

$$\begin{aligned}\hat{G}_2(u) &= \frac{1}{R_1 - V_1} \sum_{i=1}^{n} \mathbf{1}\{\xi_{i2} \le u\} \mathbf{1}\{\theta_i = 1 \text{ and } H_i \in S_1\} \\ &= \frac{1}{(R_1 - V_1)/n} \times \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\xi_{i2} \le u\} \mathbf{1}\{\theta_i = 1, \xi_{i1} \le u_1\} + o_p(1).\end{aligned}$$

By the LLN and $u_1 > 0$, it follows that

$$\hat{G}_2(u) \xrightarrow{\text{a.s.}} \frac{P(\xi_2 \le u,\, \theta = 1,\, \xi_1 \le u_1)}{P(\theta = 1,\, \xi_1 \le u_1)} = P\left(\xi_2 \le u \,|\, \xi_1 \le u_1,\, \theta = 1\right) = G_2(u).$$

Suppose we have shown that for $1 \le k < K$, as $n \to \infty$,

$$S_k = \{H_i : \xi_{is} \le u_s,\, s = 1, \dots, k\} \bigtriangleup A_k, \quad \text{with } |A_k| = o_p(n) \tag{A.1}$$

and

$$\begin{aligned}&\frac{R_k}{R_{k-1}} \xrightarrow{\text{a.s.}} \frac{u_k}{\alpha_k}, \quad \frac{R_k - V_k}{R_{k-1} - V_{k-1}} \xrightarrow{\text{a.s.}} G_k(u_k)\,, \quad \frac{V_k}{R_k} \xrightarrow{\text{a.s.}} a_k\,, \\ &\hat{G}_{k+1}(u) := \frac{1}{R_k - V_k} \sum_{i=1}^{n} \mathbf{1}\{\xi_{i,k+1} \le u,\, \theta_i = 1,\, H_i \in S_k\} \xrightarrow{\text{a.s.}} G_{k+1}(u)\,,\ \forall u.\end{aligned} \tag{A.2}$$

Since G is continuous and argument-wise strictly concave, G_{k+1} is continuous and strictly concave, which implies $\lim_{\epsilon \to 0} \tau(\alpha_{k+1}, a_k + \epsilon, G_{k+1}) = u_{k+1}$. Then by (A.2), it is not hard to see $\tau(\alpha_{k+1}, V_k/R_k, \hat{G}_k) \xrightarrow{\text{a.s.}} u_{k+1}$. By a modified argument in [11], it follows that

$$\begin{aligned}S_{k+1} &= \{H_i \in S_k : \xi_{i,k+1} \le u_{k+1}\} \bigtriangleup B_{k+1} \\ &= \{H_i : \xi_{is} \le u_s,\, s = 1, \dots, k+1\} \bigtriangleup A_{k+1}\end{aligned}$$

where $|B_{k+1}| = o_p(n)$ and $A_{k+1} = A_k \cup B_{k+1}$. By the assumption of the induction, $|A_{k+1}| = o_p(n)$ and hence (A.1) holds for S_{k+1}. Based on this, by the same argument for $k = 1$, it is not hard to show (A.2) for $k + 1$. Therefore, by induction, (A.1) and (A.2) are true for all k. The convergences in (4.10) then follow from (A.2). □

Proof of Corollary 1 By Theorem 2, as $n \to \infty$,

$$\frac{V_K}{V_0} \xrightarrow{\text{a.s.}} \prod_{k=1}^K u_k, \quad \frac{R_K - V_K}{n - V_0} \xrightarrow{\text{a.s.}} \prod_{k=1}^K G_k(u_k), \quad \frac{V_K}{R_K} \xrightarrow{\text{a.s.}} (1-a)\alpha\,. \tag{A.3}$$

By the first two limits and $V_0/n \xrightarrow{\text{a.s.}} 1-a$,

$$\frac{V_K}{R_K} \to \frac{(1-a)\prod_{k=1}^K u_k}{(1-a)\prod_{k=1}^K u_k + a\prod_{k=1}^K G_k(u_k)}\,.$$

From $G_1(u_1)\cdots G_K(u_K) = G(u_1,\ldots,u_K)$, (3.7) follows by comparing the third limit in (A.3) and the above one. □

Proof of Theorem 3. Because $(1-a+ag(\mathbf{0}))^{-1} = \alpha_* < \alpha < 1$, by the median value theorem, there is $\boldsymbol{u} = (u_1,\ldots,u_K) \in (0,1)^K$, such that

$$\frac{u_1\cdots u_K}{(1-a)u_1\cdots u_K + aG(\boldsymbol{u})} = \alpha.$$

Let $a_0 = a$. For $k \ge 1$, define G_k by (3.5) and $a_k = 1-(1-a_{k-1})\alpha_k$. Then let

$$\alpha_k = \frac{u_k}{(1-a_{k-1})u_k + a_{k-1}G_k(u_k)}\,.$$

By some algebra, it can be shown that $\alpha = \alpha_1\cdots\alpha_K$. Because each G_k is strictly concave, $\alpha_k < 1$. By the selection of α_k, it is clear that $\tau(\alpha_k, a_{k-1}, G_k) > 0$. Therefore, by Theorem 2, as $n \to \infty$, $R_K \xrightarrow{\text{a.s.}} \infty$ and $V_K/R_K \xrightarrow{\text{a.s.}} (1-a)\alpha$. By dominated convergence, $E[V_K/R_K \,|\, R_K > 0] = (1-a)\alpha + o(1)$. □

Proof of Theorem 4. Given $\alpha \in (\alpha_*, 1)$, by Theorem 3, one can assign target FDR control levels $\alpha_1,\ldots,\alpha_K$ to individual steps so that the sequential procedure attains pFDR $= (1-a)\alpha + o(1)$ with a positive power. By Corollary 1, the power is upper bounded by $P_*(\alpha)$. This proves the first half of the theorem. Because G is continuous, there is $\boldsymbol{u} \in [0,1]$ which is a solution to (3.7), such that $G(\boldsymbol{u}) = P_*(\alpha)$. Now the same construction of $\alpha_1, \ldots, \alpha_K$ as in the proof of Theorem 3 can be applied, showing that $P_*(\alpha)$ is attainable.

It is easy to see that the above argument applies to any permutation of $1, \ldots, K$ as well, and hence $P_*(\alpha)$ is attainable for any given order in which the components of the p-values are incorporated. $\square$

To prove Propositions 3 and 4, let G_k be the marginal distributions of G. In each example, for $\alpha_1, \alpha_2 \in (0, 1)$, with $\alpha_1\alpha_2 = \alpha$, the power of the sequential procedure is

$$\frac{R_2 - V_2}{n - V_0} \to G_1(u_1)\, G_2(u_2) = \frac{1}{a}\left(\frac{1}{\alpha} - 1 + a\right) u_1\, u_2\,,$$

where $u_k > 0$ satisfy

$$\frac{1}{a}\left(\frac{1}{\alpha_1} - 1 + a\right) u_1 = G_1(u_1)\,, \quad \frac{1}{a_1}\left(\frac{1}{\alpha_2} - 1 + a_1\right) u_2 = G_2(u_2)\,. \tag{A.4}$$

with $a_1 = 1 - (1 - a)\alpha_1$. We need the following lemma to show Proposition 3.

Lemma 1 The following statements are true.

(1) As $\alpha_1 \to 0$,

$$G_1(u_1) = \exp\left\{-\frac{1}{2}\left(\frac{L(\alpha_1)}{\mu_1} - \frac{\mu_1}{2}\right)^2\right\}(1 + o(1))\,. \tag{A.5}$$

(2) Fix $\epsilon > 0$. Then there is $r(\alpha_2) \to 0$ as $\alpha_2 \to 0$, such that

$$G_2(u_2) \le \exp\left\{-\frac{1}{2}\left(\frac{L(\alpha) - \delta}{\mu_2} - \frac{\mu_2}{2}\right)^2\right\}(1 + r(\alpha_2))\,, \quad \forall \alpha_1 \in [\epsilon, 1], \tag{A.6}$$

where

$$\delta = \delta(\alpha_1) = \log\left[\frac{1}{a}\left(\frac{1}{\alpha_1} - 1 + a\right)\right].$$

(3) As $(\alpha_1, \alpha_2) \to (0, 0)$, the power is

$$G_1(u_1)\, G_2(u_2) = \frac{1}{\sqrt{a\alpha}} \exp\left\{-\frac{\mu_1^2 + \mu_2^2}{8}\right\} \times \exp\left\{-\frac{1}{2}\left[\frac{L(\alpha_1)^2}{\mu_1^2} + \frac{(L(\alpha) - L(\alpha_1))^2}{\mu_2^2}\right]\right\}(1 + o(1))\,. \tag{A.7}$$

Proof of Proposition 3. Assume the Lemma is true. If $\alpha_2 \not\to 0$ as $\alpha \to 0$, then $L(\alpha_1) = L(\alpha) + O(1)$ and by (A.5), the power is

$$G_1(u_1)G_2(u_2) = O(G_1(u_1)) = O\left(\exp\left\{-\frac{L(\alpha)^2}{2\mu_1^2}\right\}\right) = o(P_*(\alpha)).$$

Likewise, if $\alpha_1 \not\to 0$, then again the power is asymptotically $o(P_*(\alpha))$. Therefore, it all remains to show that as $(\alpha_1, \alpha_2) \to (0, 0)$, the optimal power $\sim P_*(\alpha)$ and the corresponding FDR control levels are $\hat{\alpha}_1$ as stated. However, this easily follows from maximizing the quadratic function of $L(\alpha_1)$ in (A.7). □

Proof of Lemma 1. We shall apply the following standard asymptotic results

$$\Phi(x) = \frac{(1 + o(1))e^{-x^2/2}}{\sqrt{2\pi}|x|}, \quad \text{as } x \to -\infty$$

$$\Phi^*(u) = -\underbrace{\sqrt{2\log(1/u) - \log\log(1/u) - \log(4\pi) + o(1)}}_{H(u)}, \quad \text{as } u \to 0\,.$$

Then, as $u \to 0$, by $H(u) = \sqrt{2\log(1/u)} + o(1)$, for $k = 1, 2$,

$$\begin{aligned}
G_k(u) = \Phi(\Phi^*(u) + \mu_k) &= \frac{e^{-\frac{1}{2}(H(u)-\mu_k)^2}}{\sqrt{2\pi}(H(u) - \mu_k)}(1 + o(1)) \\
&= \exp\left\{-\log\frac{1}{u} + \frac{1}{2}\log\log\frac{1}{u} + \frac{1}{2}\log(4\pi) + \mu_k\sqrt{2\log\frac{1}{u}} - \frac{1}{2}\mu_k^2\right\}\frac{1 + o(1)}{\sqrt{4\pi\log(1/u)}} \\
&= u\exp\left\{\mu_k\sqrt{2\log(1/u)} - \frac{1}{2}\mu_k^2\right\}(1 + o(1))
\end{aligned}$$

By (A.4), $e^{\delta}u_1 = G_1(u_1)$ and $e^{\Delta-\delta}u_2 = G_2(u_2)$, where

$$\Delta = \log\left[\frac{1}{a}\left(\frac{1}{\alpha} - 1 + a\right)\right] \implies \Delta - \delta = \log\left[\frac{1}{a_1}\left(\frac{1}{\alpha_2} - 1 + a_1\right)\right].$$

(1) Let $\alpha_1 \to 0$. Then $u_1 = e^{-\delta}G_1(u_1) \to 0$ and hence by the asymptotics of $G_1(u)$,

$$\begin{aligned}
e^{\delta}u_1 = G_1(u_1) &= u_1\,\exp\left\{\mu_1\sqrt{2\log\frac{1}{u_1}} - \frac{1}{2}\mu_1^2\right\}(1 + o(1)) \\
\implies u_1 &= \exp\left\{-\frac{1}{2}\left(\frac{\delta}{\mu_1} + \frac{\mu_1}{2}\right)^2 + o(1)\right\} = \exp\left\{-\frac{1}{2}\left(\frac{\delta}{\mu_1} + \frac{\mu_1}{2}\right)^2\right\}(1 + o(1)).
\end{aligned}$$

By $e^\delta u_1 = G_1(u_1)$ again,

$$G_1(u_1) = \exp\left\{-\frac{1}{2}\left(\frac{\delta}{\mu_1} - \frac{\mu_1}{2}\right)^2\right\}(1 + o(1))$$

Notice that when $\alpha_1 \to 0$, $\delta = L(\alpha_1) + o(1)$. Then (A.5) follows.

(2) Fix $\epsilon > 0$. As $\alpha_2 \to 0$, $u_2 = e^{-\Delta+\delta} G_2(u_2) \to 0$ uniformly for all $\alpha_1 \geq \epsilon$. Then

$$G_2(u_2) = \exp\left\{-\frac{1}{2}\left(\frac{\Delta - \delta}{\mu_2} - \frac{\mu_2}{2}\right)^2\right\}(1 + r),$$

where $r = r(\alpha_1, \alpha_2)$ with $\sup_{\alpha_1 \geq \epsilon} |r| \to 0$ as $\alpha_2 \to 0$. By $\Delta = L(\alpha) + o(1)$, (A.6) follows.

(3) Finally, as $(\alpha_1, \alpha_2) \to (0, 0)$, $a_1 \to 1$, $\delta \to \infty$, and $\Delta - \delta \to \infty$. As a result, $u_k \to 0$ for $k = 1, 2$, and so from the above asymptotics,

$$G_1(u_1)G_2(u_2) = \exp\left\{-\frac{1}{2}\left[\frac{\delta^2}{\mu_1^2} + \frac{(\Delta - \delta)^2}{\mu_2^2} - \Delta + \frac{\mu_1^2 + \mu_2^2}{4}\right] + o(1)\right\}$$

Note that $\delta = L(\alpha_1) + o(1)$ and $\Delta = L(\alpha) + o(1)$. The proof is then complete. □

Proof of Proposition 4. Let $\nu_k > 1$ be the conjugates of μ_k, i.e., $1/\mu_k + 1/\nu_k = 1$. Then $G_k(u) = u^{1-1/\nu_k}$. Let $m = \frac{1}{a}(\frac{1}{\alpha_1} - 1 + a)$. Then by (A.4), the power is asymptotically equal to $\rho(\alpha) u_1 u_2$, where u_k solve the equations $m u_1 = u_1^{1/\mu_1}$, and $[\rho(\alpha)/m] u_2 = u_2^{1/\mu_2}$. Therefore, by Corollary 1, the power asymptotically is equal to

$$\rho(\alpha) u_1 u_2 = \rho(\alpha) m^{-\nu_1} [\rho(\alpha)/m]^{-\nu_2} = \rho(\alpha)^{1-\nu_2} m^{\nu_2 - \nu_1}.$$

If $\mu_1 = \mu_2$, then $\nu_1 = \nu_2$ and hence for any α_1, α_2, the power is asymptotically equal to $\rho(\alpha)^{1-\nu_2} = \rho(\alpha)^{-1/(\mu_2 - 1)} = \rho(\alpha)^{-1/c}$. If $\mu_1 > \mu_2$, then $\nu_1 < \nu_2$. Because $\rho(\alpha) \geq m > 1$, $m^{\nu_2 - \nu_1} \leq \rho(\alpha)^{\nu_2 - \nu_1}$, with "=" iff $(\alpha_1, \alpha_2) = (\alpha, 1)$ and the maximum power is $\rho(\alpha)^{1-\nu_1} = \rho(\alpha)^{-1/(\mu_1 - 1)} = \rho(\alpha)^{-1/c}$. When $\mu_1 < \mu_2$, then $\nu_1 > \nu_2$ and hence $m^{\nu_2 - \nu_1} \leq 1$ with "=" iff $(\alpha_1, \alpha_2) = (1, \alpha)$. In this case, the maximum power is $\rho(\alpha)^{1-\nu_2} = \rho(\alpha)^{-1/c}$. □

A.2 Technical details for the simultaneous procedure

Proof of Proposition 5. First, it is not difficult to see that the function

$$\phi_k(\boldsymbol{t}) := f_k\left(\frac{R(\boldsymbol{t}) \vee 1}{n}\right), \quad k = 1, \ldots, K$$

are right-continuous, i.e., as $\boldsymbol{s} \to \boldsymbol{t}+$, $\phi_k(\boldsymbol{s}) \to \phi_k(\boldsymbol{t})$. Suppose $\boldsymbol{t} \in \sup D$. Then for any $\boldsymbol{t}' > \boldsymbol{t}$, $t'_k \geq \alpha_k \phi_k(\boldsymbol{t}')$. Letting $\boldsymbol{t}' \to \boldsymbol{t}$ yields $t_k \geq \alpha_k \phi_k(\boldsymbol{t})$. On the other hand, there is a sequence $\hat{\boldsymbol{t}}_n < \boldsymbol{t}$ and $\hat{\boldsymbol{t}}_n \to \boldsymbol{t}$, such that $\hat{t}_{n,k} \leq \alpha_k \phi_k(\hat{\boldsymbol{t}}) \leq \alpha_k \phi_k(\boldsymbol{t})$. Letting $n \to \infty$ gives $t_k \leq \alpha_k \phi_k(\boldsymbol{t})$. Therefore,

$$t_k = \alpha_k \phi_k(\boldsymbol{t}), \quad k = 1, \ldots, K \tag{A.8}$$

Suppose $\sup D$ contains two points $\boldsymbol{s} \neq \boldsymbol{t}$. Without loss of generality, assume $t_1 < s_1$. Then by (A.8), $\phi_1(\boldsymbol{t}) < \phi_1(\boldsymbol{s})$. Because f_1 is nondecreasing, by its definition, $R(\boldsymbol{t}) < R(\boldsymbol{s})$. As a result, $t_k = \phi_k(t) \leq \phi_k(\boldsymbol{s}) = s_k$ for all k. Since $t_1 < s_1$, then $\boldsymbol{t} < \boldsymbol{s}$, which contradicts the definition of $\sup D$. Therefore, $\sup D$ has only one element $\boldsymbol{t}_0 = (t_1, \ldots, t_K)$.

To show that the procedures in (4.4) and (4.6) are equivalent, let $\tau' = (R(\boldsymbol{t}_0) \vee 1)/n$. Then $\tau' \in [0, 1]$ and by (A.8), $t_k = \alpha_k f_k(\tau')$. Then $\boldsymbol{t}_0 = \boldsymbol{\gamma}(\tau')$ and $\tau' = (R(\boldsymbol{\gamma}(\tau')) \vee 1)/n$. Clearly, $\tau' \in S := \{s \in [0, 1] : s \leq (R(\boldsymbol{\gamma}(s)) \vee 1)/n\}$. On the other hand, for any $s \in S$, $\boldsymbol{\gamma}(s) \in D$. Therefore, $\boldsymbol{\gamma}(s) \leq \boldsymbol{t}_0 = \boldsymbol{\gamma}(\tau')$, i.e., $f_k(s) \leq f_k(\tau')$. Take product over k. By $f_1(t) \cdots f_K(t) = t$, $s \leq \tau'$. As a result, $\tau' = \tau = \sup S$. This proves the equivalence of the two procedures. It is easy to see that the procedures in (4.4) and (4.5) are equivalent. For brevity, the detail of proof is omitted. □

The proof gives the following result, which will be used in the proof for Theorem 5,

$$\tau = \frac{R(\boldsymbol{\gamma}(\tau)) \vee 1}{n} \tag{A.9}$$

Proof of Theorem 5. The proof is based on Proposition 5 as well as the following Lemma, which obtains certain martingale structure from the p-values.

Lemma 2 Let $V(\boldsymbol{s})$ be the total number of true nulls H_i with p-values $\boldsymbol{\xi}_i \leq \boldsymbol{s}$. Then the process

$$X(t) := \frac{V(\boldsymbol{\gamma}(t))}{t}, \quad t \in [0,1],$$

is a martingale running backward in time with respect to $\mathcal{F}_t = \sigma(\mathbf{1}\left\{\boldsymbol{\xi}_j \leq \boldsymbol{\gamma}(s)\right\}, t \leq s \leq 1,\ j = 1, \ldots, n)$ and τ is a stopping time with respect to $\mathcal{F}_t$.

Proof of Theorem 5. Assume the lemma is true for now. By $V = V(\boldsymbol{t}_0)$ and $\boldsymbol{t}_0 = \boldsymbol{\gamma}(\tau)$, we have $V = V(\boldsymbol{\gamma}(\tau))$. Then by Lemma 2 and the optional stopping theorem,

$$E\left[\frac{V}{n\tau}\right] = E\left[\frac{V(\boldsymbol{\gamma}(\tau))}{n\tau}\right] = \frac{1}{n}E[V(\boldsymbol{\gamma}(1))] = \frac{1}{n}E[V(\boldsymbol{\alpha})] = (1-a)\prod_{k=1}^{K}\alpha_k = (1-a)\alpha.$$

On the other hand, $R = R(\boldsymbol{t}_0) = R(\boldsymbol{\gamma}(\tau))$. By (A.9), $n\tau = R(\boldsymbol{\gamma}(\tau)) \vee 1 = R \vee 1$. This combined with the above equation then proves the statement. □

Proof of Lemma 2. Since the p-values of true nulls are iid $\sim \text{Unif}(0,1)^{\otimes K}$, by $f_1(t)\cdots f_K(t) = t$, for any $s < t$, $E[V(\boldsymbol{\gamma}(s)) \,|\, V(\boldsymbol{\gamma}(t))] = \frac{s}{t}V(\boldsymbol{\gamma}(t))$. Therefore, $X(t)$ is a martingale running backward in time. □

Proof of Proposition 6. By the LLN, for every $s \in [0,1]$,

$$\frac{R(\boldsymbol{\gamma}(s)) \vee 1}{n} \xrightarrow{\text{a.s.}} (1-a)\prod_{k=1}^{K}(\alpha_k f_k(s)) + a\,G(\boldsymbol{\gamma}(s)), \quad \text{as } n \to \infty.$$

Since $f_1(s)\cdots f_K(s) = s$, the right hand side is $h(s)$. On the other hand, by Proposition 5,

$$\tau = \sup\left\{s \in [0,1]:\ s \leq \frac{R(\boldsymbol{\gamma}(s)) \vee 1}{n}\right\}$$

Following [11], it then can be shown that $\tau \to s^*$ in probability. The argument can be modified to show that $\tau \to s^*$ with probability 1 [cf. 6]. For brevity, the detail is omitted.

Suppose the conditions in (4.9) are satisfied. Then

$$\frac{h(s)}{s} = (1-a)\alpha + \frac{a\,G(\boldsymbol{\gamma}(s))}{\prod_{k=1}^K f_k(s)}$$

Since $f_k(s) > 0$ for all $s > 0$ and g is continuous, $f_1(s) \ldots f_K(s) = s$,

$$\frac{G(\boldsymbol{\gamma}(s))}{s} = \frac{G(\alpha_1 f_1(s), \ldots, \alpha_K f_K(s))}{\prod_{k=1}^K f_k(s)} \to g(\mathbf{0}) \prod_{k=1}^K \alpha_k = \alpha g(\mathbf{0}), \quad s \to 0+.$$

Therefore, $h(s)/s \to \alpha\,(1 - a + a g(\mathbf{0})) > 1$. As a result, $h(s) > s$ for all $s > 0$ small enough. Since $h(1) < 1$ and h is continuous, $h(s) = s$ has a positive solution in $(0, 1)$.

□

The proof of Theorem 6 follows from Proposition 6 and the LLN. Since it is similar to the one for Theorem 2, the detail is omitted for brevity.

Proof of Theorem 7. By Proposition 5 and Theorem 6, $D \neq \emptyset$ and $P_*(\alpha) > 0$. From Corollary 2, $P_*(\alpha)$ is an upper bound for the power of the simultaneous procedure with the target FDR control level α. Since $\alpha < 1$, by (4.11), for each $\boldsymbol{u} \in D$, $\prod_{k=1}^K u_k = G(\boldsymbol{u})/\rho(\alpha) < 1$, yielding $\boldsymbol{u} < \mathbf{1}$. It is easy to see that D is compact. Let $\boldsymbol{v} \in D$ such that $P_*(\alpha) = G(\boldsymbol{v})$. Note that each $v_k > 0$ and it is possible that $v_k = 1$ for some but not all k. If $P_*(\alpha) = 1$, then by (4.11), $\rho(\alpha) \prod_{k=1}^K v_k = 1$. Since $v_k \le 1$, it can be seen $\alpha = 1$, which is a contradiction. Therefore, $P_*(\alpha) \in (0, 1)$.

To show the second part of the result, from $\rho(\alpha) \prod_{k=1}^K v_k < 1$,

$$v_1 \cdots v_K < \frac{\alpha a}{1 - \alpha + \alpha a} < \alpha.$$

Therefore, there are $0 < \alpha_1, \ldots, \alpha_K \le 1$ such that (1) $\alpha = \alpha_1 \cdots \alpha_K$, (2) for $v_k < 1$, $\alpha_k \in (v_k, 1)$ and (3) for $v_k = 1$, $\alpha_k = 1$. Let $s^* = v_1 \cdots v_K/\alpha$. Then $s^* \in (0, 1)$.

Now for each k with $v_k = 1$, define $f_k(t) = t^{q_k}$, with $q_k = \log(v_k/\alpha_k)/\log s^*$. It is easy to see $q_1 + \cdots + q_K = 1$. For $v_k = 1$, $q_k = 0$; and for $v_k < 1$, since $v_k < \alpha_k$, $q_k > 0$. Consider the simultaneous procedure based on α_k and f_k. Define $h(s)$ according to (4.7). Because $\rho(\alpha)\prod_{k=1}^K v_k = G(\boldsymbol{v})$ and $\alpha_k f_k(s^*) = v_k$, s^* is a solution to $s = h(s)$. Therefore, by Theorem 6, the power of procedure has limit equal to $\alpha\rho(\alpha)s^* = \rho(\alpha)v_1 \cdots v_K = G(\boldsymbol{v}) = P_*(\alpha)$. The proof is thus complete. □

Proof of Proposition 7. By Theorems 4 and 7, the maximum power of a sequential or simultaneous procedure with the target FDR control level α is $G(\boldsymbol{u})$ for some $\boldsymbol{u} \in [0,1]^K$ that solves $\rho(\alpha)\prod_{k=1}^K u_k = G(\boldsymbol{u})$. If there are more than one $u_k < 1$, then by (4.12), there is k such that $G(\boldsymbol{u}) < G_k(u_1 \cdots u_K)$. Then the equation $\rho(\alpha)u = G_k(u)$ has a solution $v > u_1 \cdots u_K$. As a result, the BH procedure using $\xi_{1k}, \ldots, \xi_{nk}$ alone has power $G_k(v) = \rho(\alpha)v > \rho u_1 \cdots u_K = G(\boldsymbol{u})$. □

Proof of Proposition 8. Assume that the statement is not true and, without loss of generality, as $\alpha \to \alpha_*$, $(u_1, \ldots, u_j) \to 0$ but $u_k \to u_k^* > 0$ for $k > j$, where $j < K$. From $\rho(\alpha) = G(\boldsymbol{u})/\prod_{k=1}^K u_k$ and $\rho(\alpha) \to g(\mathbf{0})$, it follows

$$g(\mathbf{0}) = \frac{1}{u_j \cdots u_K} \int_{\xi_k \le u_k^*,\ k>j} g(0, \ldots, 0, \xi_{j+1}, \ldots, \xi_K)\, d\xi_{j+1} \cdots d\xi_K.$$

which is impossible because $g(\boldsymbol{x}) < g(\mathbf{0})$ for all $\boldsymbol{x} \ne \mathbf{0}$.

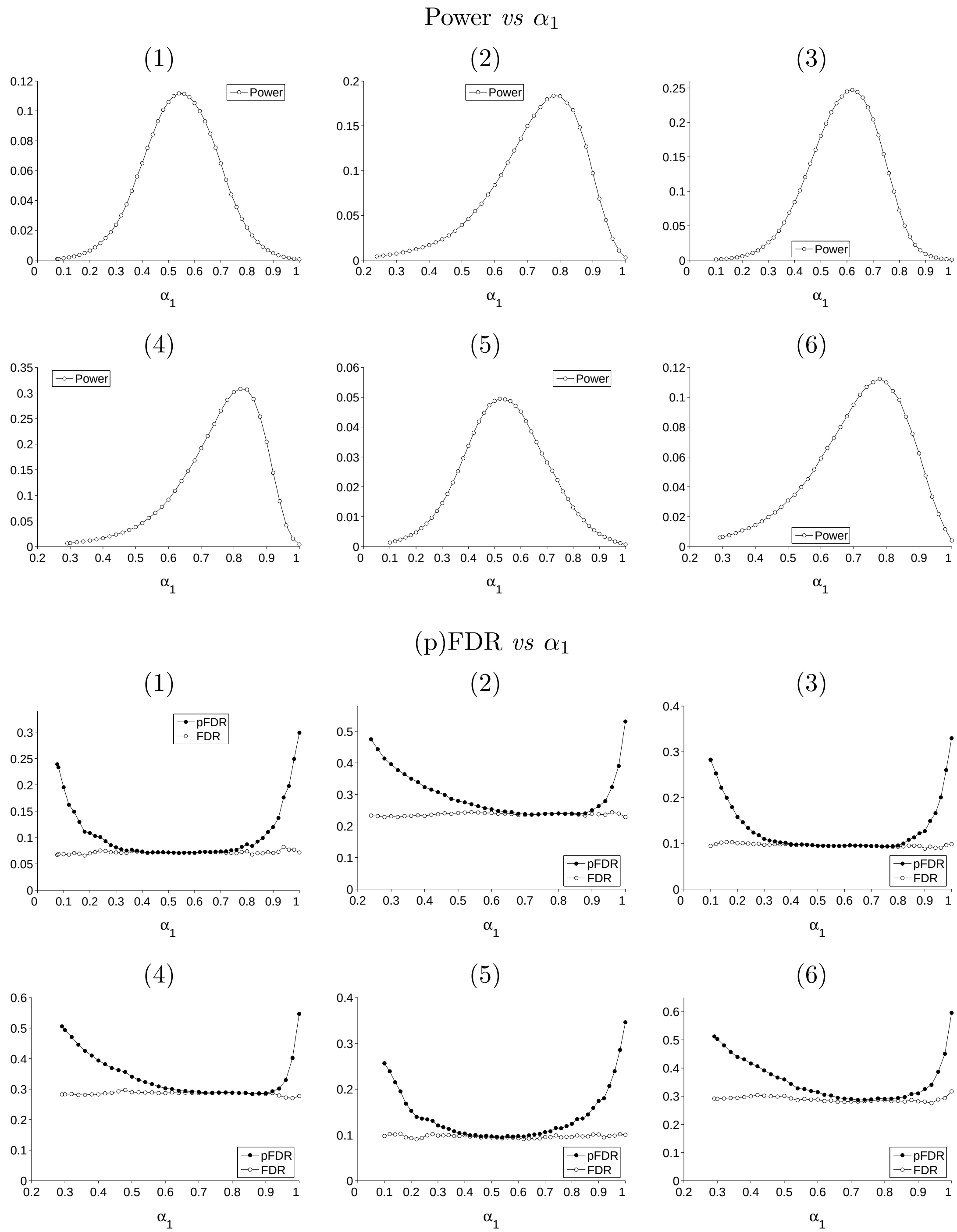


Figure 5.1: Sequential procedure involving t-distributions: power *vs* α_1 (top) and (p)FDR *vs* α_1 (bottom), where α_1 is the target FDR control level of the first step. Each plot corresponds to a combination of a, Σ_a, and α displayed in Table 5.1.

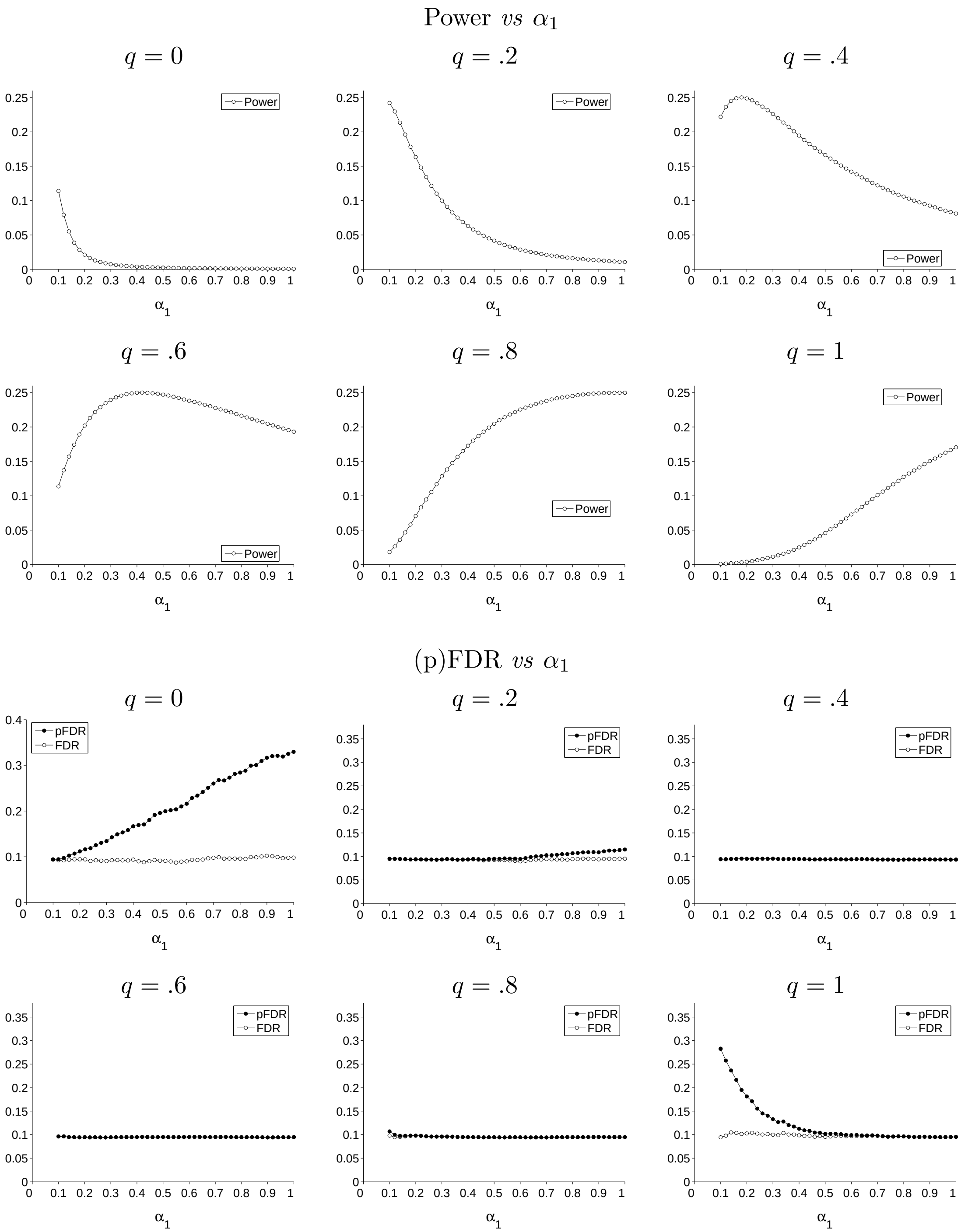


Figure 5.2: Simultaneous procedure involving t-distributions. Each plot is associated with a different $f_1(x) = x^q$. For all the plots, $a = .5$, $\Sigma_a = AA^T$. The procedure is applied to the same randomly sampled p-values as the sequential procedure in plots (3) of Fig. 5.1.

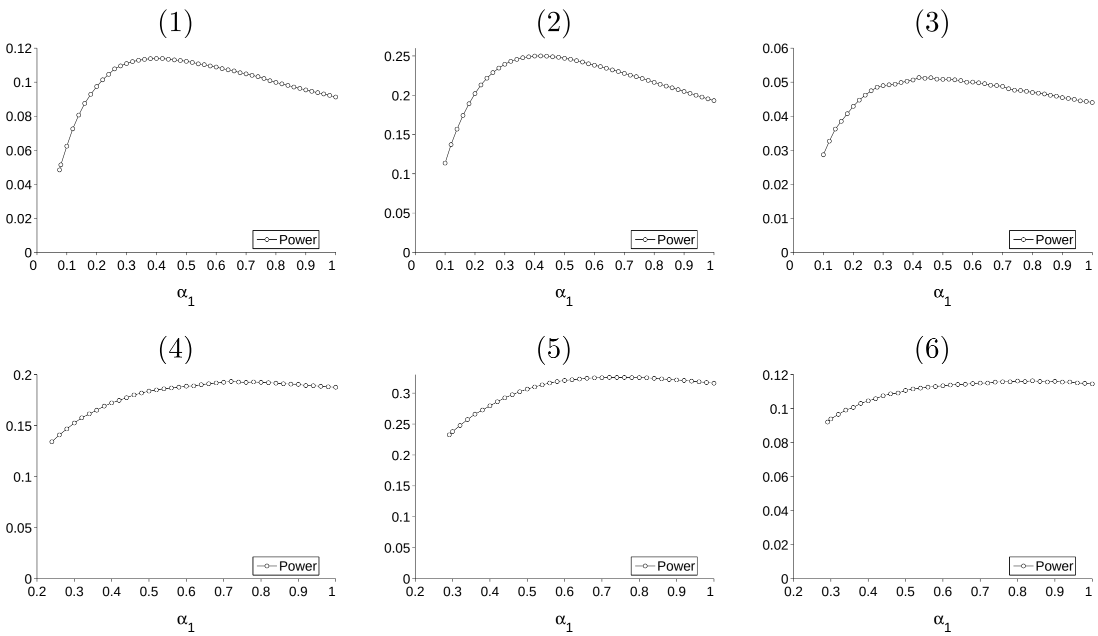


Figure 5.3: Simultaneous procedure involving t-distributions: power *vs* α_1 associated with $f_1(x) = x^{0.6}$, but with different combinations of a, Σ_a, and α as in Table 5.1.

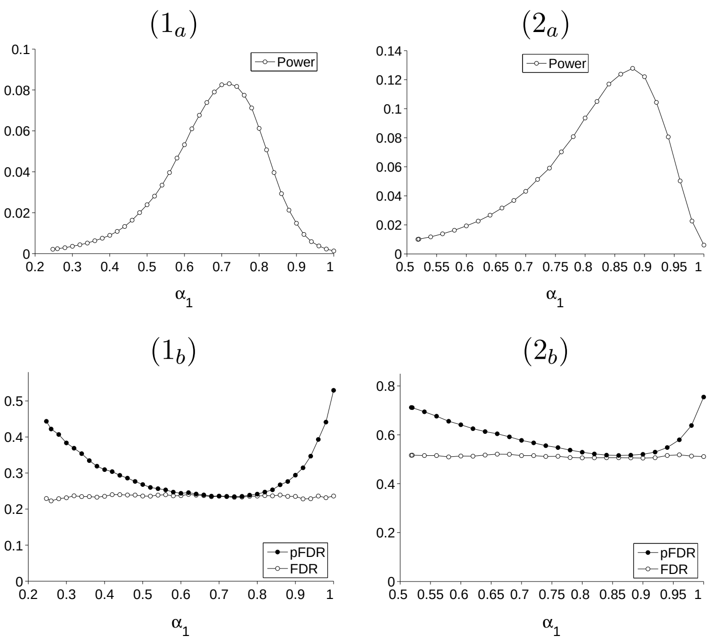


Figure 5.4: Sequential procedure involving F-distributions: power *vs* α_1 (top) and (p)FDR *vs* α_1 (bottom), where α_1 is the target FDR control level of the first step. Each column of plots correspond to a pair (a, α) displayed in Table 5.2.

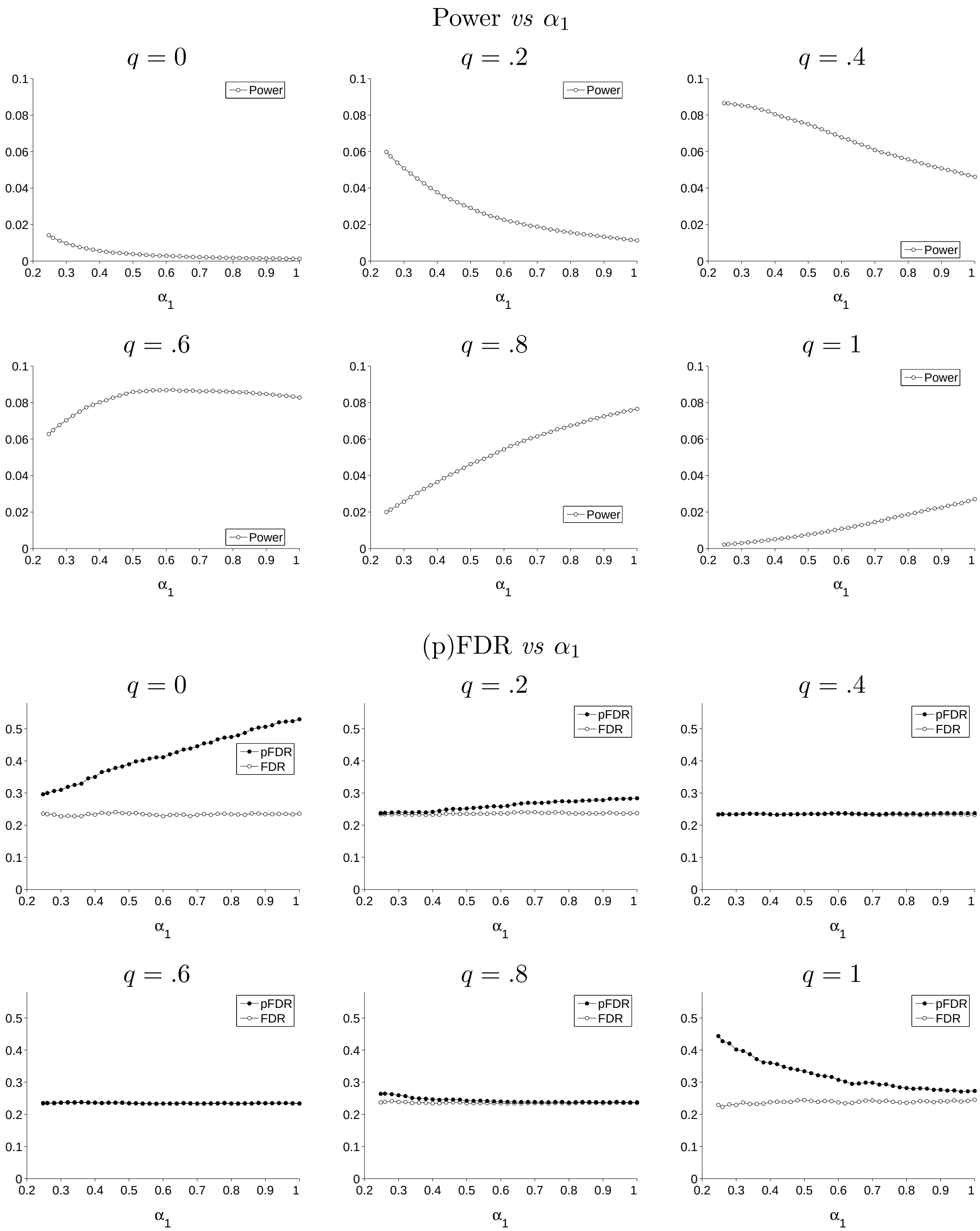


Figure 5.5: Simultaneous procedure involving F-distributions: power *vs* α_1 (top) and (p)FDR *vs* α_1 associated with different $f_1(x) = x^q$. The procedure is applied to the same randomly sampled p-values as the sequential procedure in plots (1) of Fig. 5.4. The plots correspond to the column $a = .05$ in Table 5.2.

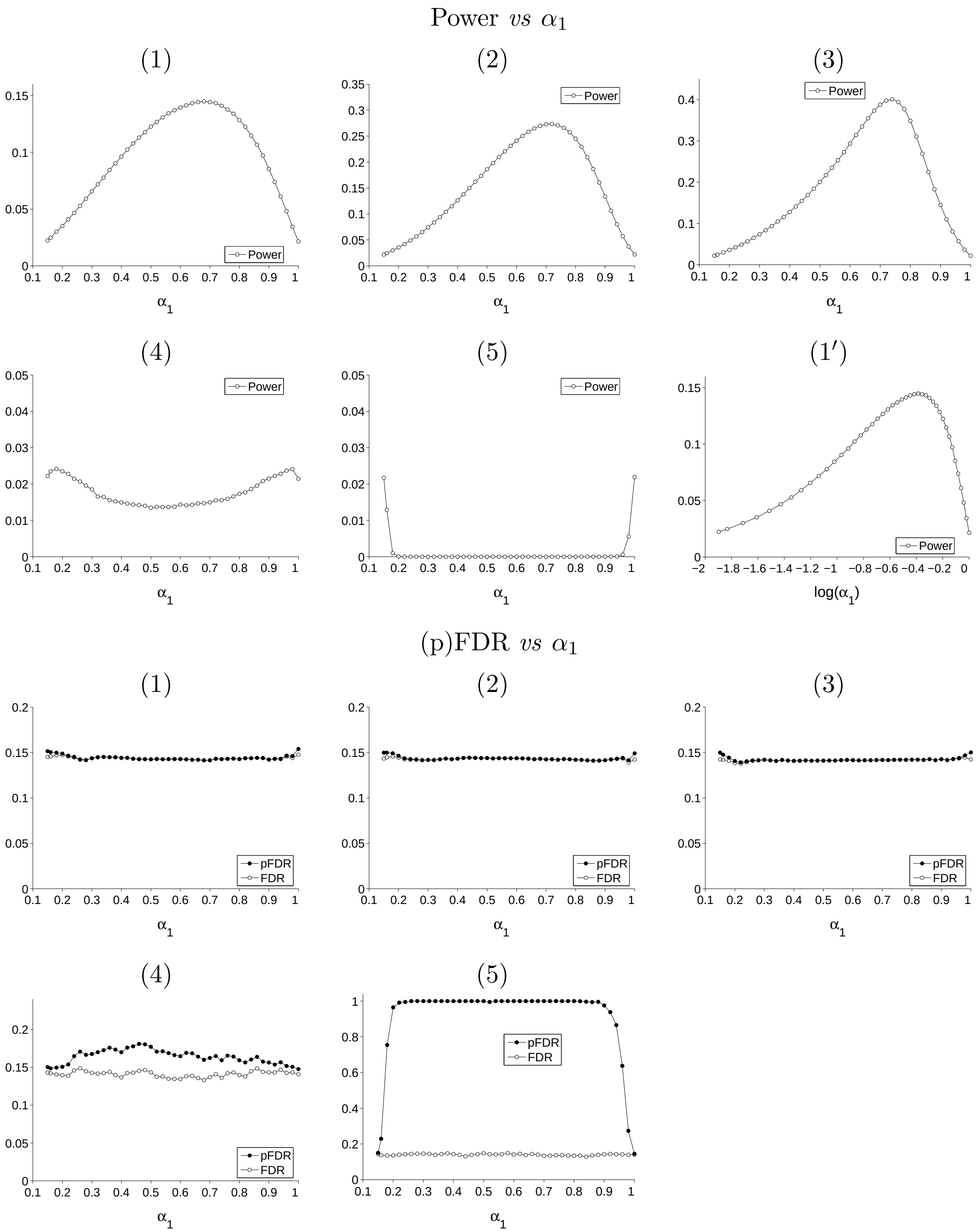


Figure 5.6: Sequential procedure involving Normal distributions: power *vs* α_1 (top, plots 1–5) and (p)FDR *vs* α_1 (bottom). In each panel, plots (1)–(5) correspond to $\rho = 0, .5, .9, -.5$ and $-.9$. For all the plots, $a = .05$. Plot $(1')$ in the top panel is the same as plot (1), except that the x-axis corresponds to $\log(\alpha_1)$.